ENCYCLOPEDIA OF

Physical Science and Technology

THIRD EDITION

Biochemistry

AP

Table of Contents
(Subject Area: Biochemistry)

Article	*Authors*	Pages in the Encyclopedia
Bioenergetics	*Richard E. McCarty and Eric A. Johnson*	Pages 99-115
Enzyme Mechanisms	*Stephen J. Benkovic and Ann M. Valentine*	Pages 627-639
Food Colors	*Pericles Markakis*	Pages 105-120
Glycoconjugates and Carbohydrates	*Eugene A. Davidson*	Pages 833-849
Ion Transport Across Biological Membranes	*George P. Hess*	Pages 99-108
Lipoprotein/Cholesterol Metabolism	*Alan D. Attie*	Pages 643-660
Membrane Structure	*Anna Seelig and Joachim Seelig*	Pages 355-367
Natural Antioxidants In Foods	*Eric A. Decker*	Pages 335-342
Nucleic Acid Synthesis	*Sankar Mitra_ Tapas K. Hazra and Tadahide*	Pages 853-876
Protein Folding	*Maurice Eftink and Susan Pedigo*	Pages 179-190
Protein Structure	*Ivan Rayment*	Pages 191-218
Protein Synthesis	*Paul Schimmel and Rebecca W. Alexander*	Pages 219-240
Vitamins and Coenzymes	*David E. Metzler*	Pages 509-528

Bioenergetics

Richard E. McCarty
Eric A. Johnson
Johns Hopkins University

I. Catabolic Metabolism: The Synthesis of ATP
II. Photosynthesis
III. Origin of Mitochondria and Chloroplasts
IV. Illustrations of the Uses of ATP: Ion Transport, Biosynthesis, and Motility
V. Concluding Statements

GLOSSARY

Adenosine 5′-triphosphate (ATP) The carrier of free energy in cells.

Bioenergetics The study of energy relationships in living systems.

Chloroplasts The sites of photosynthesis in green plants.

Ion transport The movement of ions across biological membranes.

Metabolism The total of all reactions that occur in cells. Catabolic metabolism is generally degradative and exergonic, whereas anabolic metabolism is synthetic and requires energy.

Mitochondria Sites of oxidative (catabolic) metabolism in cells.

Photosynthesis Light-driven synthesis of organic molecules from carbon dioxide and water.

Plasma membrane The barrier between the inside of cells and the external medium.

BIOENERGETICS, an amalgamation of the term *biological energetics*, is the branch of biology and biochemistry that is concerned with how organisms extract energy from their environment and with how energy is used to fuel the myriad of life's endergonic processes. Organisms may be usefully divided into two broad groups with respect to how they satisfy their need for energy. Autotrophic organisms convert energy from nonorganic sources such as light or from the oxidation of inorganic molecules to chemical energy. As heterotrophic organisms, animals must ingest and break down complex organic molecules to provide the energy for life.

Interconversions of forms of energy are commonplace in the biological world. In photosynthesis, the electromagnetic energy of light is converted to chemical energy, largely in the form of carbohydrates, with high overall efficiency. The energy of light is used to drive oxidation–reduction reactions that could not take place in the dark. Light energy also powers the generation of a proton electrochemical potential across the green photosynthetic

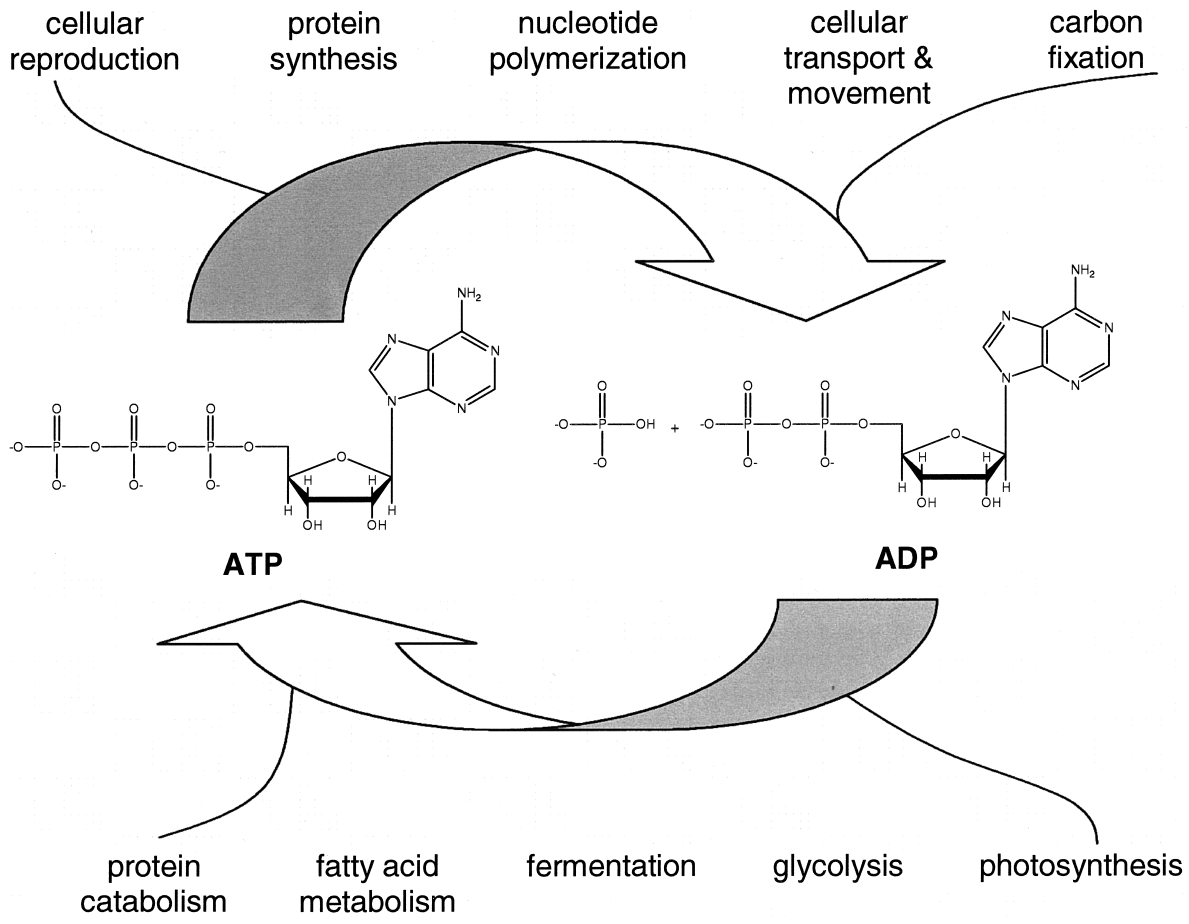


FIGURE 1 Central role of adenosine 5′-triphosphate (ATP) in metabolism. Catabolic (degradative) metabolism is exergonic and provides the energy needed for the synthesis of ATP from adenosine 5′-diphosphate (ADP) and inorganic phosphate (P_i). The exergonic hydrolysis of ATP in turn powers the endergonic processes of organisms.

membrane. Thus, electrical work is an integral part of photosynthesis. Chemical energy is used in all organisms to drive the synthesis of large and small molecules, motility at the microscopic and macroscopic levels, the generation of electrochemical potentials of ions across cellular membranes, and even light emission as in fireflies.

Given the diversity in the forms of life, it might be expected that organisms have evolved many mechanisms to deal with their need for energy. To some extent this expectation is the case, especially for organisms that live in extreme environments. However, the similarities among organisms in their bioenergetic mechanisms are as, or even more, striking than the differences. For example, the sugar glucose is catabolized (broken down) by a pathway that is the same in the enteric bacterium *Escherichia coli* as it is in higher organisms. All organisms use adenosine 5′-triphosphate (ATP) as a central intermediate in energy metabolism. ATP acts in a way as a currency of free energy. The synthesis of ATP from adenosine 5′-diphosphate (ADP) and inorganic phosphate (P_i) is a strongly endergonic reaction that is coupled to exergonic reactions such as the breakdown of glucose. ATP hydrolysis in turn powers many of life's processes. The central role of ATP in bioenergetics is illustrated in Fig. 1. Partial structures of several compounds that play important roles in metabolism are shown in Fig. 2.

In this article, the elements of energy metabolism will be discussed with emphasis on how organisms satisfy their energetic requirements and on how ATP hydrolysis drives otherwise unfavorable reactions.

I. CATABOLIC METABOLISM: THE SYNTHESIS OF ATP

Metabolism may be defined as the total of all the chemical reactions that occur in organisms. Green plants can synthesize all the thousands of compounds they contain

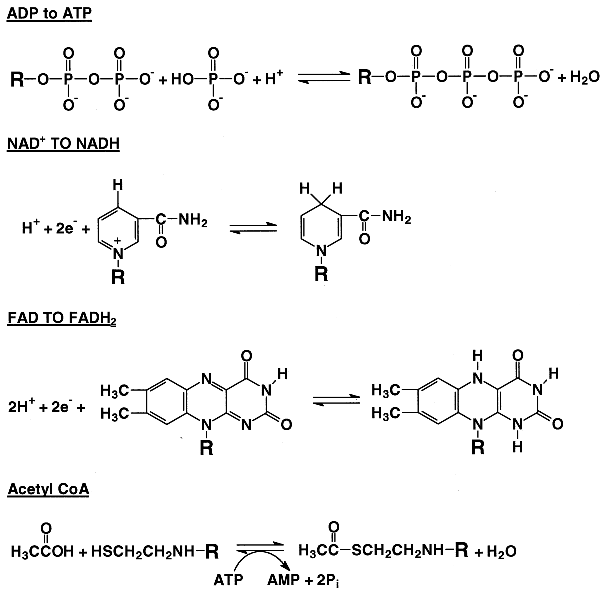


FIGURE 2 Some important reactions in metabolism. Shown are the phosphorylation of ADP to ATP, NAD^+, NADH, FAD, $FADH_2$ acetate, CoA, and acetyl CoA. For clarity, just the parts of the larger molecules that undergo reaction are shown. NAD^+, nicotinamide adenine dinucleotide; NADH, nicotinamide adenine dinucleotide (reduced form); FAD, flavin adenine dinucleotide; $FADH_2$, flavin adenine dinucleotide (reduced form); CoA, coenzyme A; AMP, adenosine monophosphate.

from carbon dioxide, water, and inorganic nutrients. The discussion of the complicated topic of metabolism is somewhat simplified by separation of the subject into two areas—catabolic and anabolic metabolism. Catabolic metabolism is degradative and is generally exergonic. ATP is a product of catabolic metabolism. In contrast, anabolic metabolism is synthetic and requires ATP. Fortunately, there are relatively few major pathways of energy metabolism.

A. Glycolysis and Fermentation

Carbohydrates are a major source of energy for organisms. The major pathway by which carbohydrates are degraded is called glycolysis. Starch, glycogen, and other carbohydrates are converted to the sugar glucose by pathways that will not be considered here. In glycolysis, glucose, a six-carbon sugar, is oxidized and cleaved by enzymes in the cytoplasm of cells to form two molecules of pyruvate, a three-carbon compound (see Figs. 3 and 4). The overall reaction is exergonic and some of the energy released is conserved by coupling the synthesis of ATP to glycolysis.

Before it may be metabolized, glucose must first be phosphorylated on the hydroxyl residue at position 6. Under intracellular conditions, the direct phosphorylation of glucose by P_i is an unfavorable reaction, characterized by a $\Delta G_0'$ of about 4 kcal/mol, at pH 7.0 and 25°C. (Note that the biochemist's standard state differs from that as usually defined in that the activity of the hydrogen ion is taken as 10^{-7} M, or pH 7.0, rather than 1 M, or pH 0.0. pH 7.0 is much closer to the pH in most cells.) This problem is neatly solved in cells by using ATP, rather than P_i, as the phosphoryl donor:

$$\text{Glucose} + \text{ATP} \longleftrightarrow \text{Glucose 6-phosphate} + \text{ADP}.$$

The $\Delta G_0'$ for this reaction, which is catalyzed by the enzyme hexokinase, is approximately −4 kcal/mol. Thus the phosphorylation of glucose by ATP is an energetically favorable reaction and is one example of how the chemical energy of ATP may be used to drive otherwise unfavorable reactions.

Glucose 6-phosphate is then isomerized to form fructose 6-phosphate, which in turn is phosphorylated by ATP at the 1-position to form fructose 1,6-bisphosphate. It seems odd that a metabolic pathway invests 2 mol of ATP in the initial steps of the pathway when ATP is an important product of the pathway. However, this investment pays off in later steps.

Fructose 1,6-bisphosphate is cleaved to form two triose phosphates that are readily interconvertible. Note that the oxidation–reduction state of the triose phosphates is the same as that of glucose 6-phosphate and the fructose phosphates. All molecules are phosphorylated sugars. In the next step of glycolysis, glyceraldehyde 3-phosphate is oxidized and phosphorylated to form a sugar acid that contains a phosphoryl group at positions 1 and 3. The oxidizing agent, nicotinamide adenine dinucleotide (NAD^+), is a weak oxidant (E_0', at pH 7.0 of −340 mV). The oxidation of the aldehyde group of glyceraldehyde 3-phosphate to a carboxylate is a favorable reaction that drives both the oxidation and the phosphorylation. This is the only oxidation–reduction reaction in glycolysis.

The hydrolysis of acyl phosphates, such as that of position 1 of 1,3-bisphosphoglycerate, is characterized by strongly negative $\Delta G_0'$ values. That for 1,3-bisphosphoglycerate is approximately −10 kcal/mol, which is significantly more negative than the $\Delta G_0'$ for the hydrolysis of ATP to ADP and P_i. Thus, the transfer of the acyl phosphate from 1,3-bisphosphoglycerate to ADP to form 3-phosphoglycerate and ATP is a spontaneous reaction. Since two sugar acid bisphosphates are formed per glucose metabolized, the two ATP invested in the beginning of the pathway have been recovered.

In the next steps of glycolysis, the phosphate on the 3-position of the 3-phosphoglycerate is transferred to the hydroxyl residue at position 2. Removal of the elements of water from 2-phosphoglycerate results in the formation of an enolic phosphate compound, phospho(enol)pyruvate

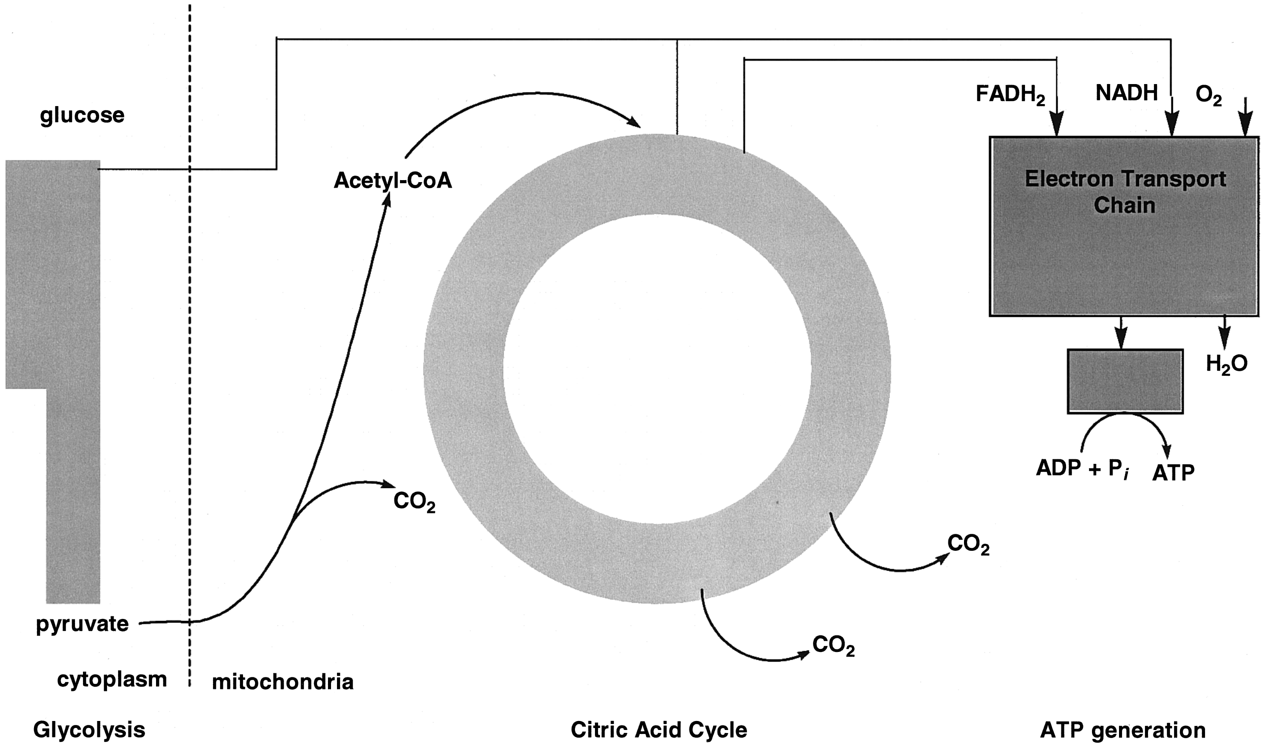


FIGURE 3 Schematic outline of carbohydrate metabolism. Glucose is oxidized to two molecules of pyruvate by glycolysis in the cytoplasm. In mitochondria, pyruvate is oxidized by molecular oxygen to CO_2 and water. The synthesis of ATP is coupled to pyruvate oxidation.

(PEP). The free energy of hydrolysis of PEP to form the enol form of pyruvate and P_i is on the order of −4 kcal/mol. In aqueous solution, however, the enol form of pyruvate is very unstable. Thus, the hydrolysis of PEP to form pyruvate is a very exergonic reaction. The $\Delta G_0'$ for this reaction is −14.7 kcal/mol, which corresponds to an equilibrium constant of 6.4×10^{10}. PEP is thus an excellent phosphoryl donor and the formation of pyruvate is coupled to ATP synthesis. Since two molecules of pyruvate are formed per glucose catabolized, two ATP are formed. Thus the net yield of ATP is two per glucose oxidized to pyruvate.

In some organisms, glycolysis is the only source of ATP. A familiar example is yeast growing under anaerobic (no oxygen) conditions. In this case, glucose is said to be fermented and ethyl alcohol and carbon dioxide (CO_2) are the end products (Fig. 5). In contrast, all higher organisms can completely oxidize pyruvate to CO_2 and water, using molecular oxygen as the terminal electron acceptor. The conversion of glucose to pyruvate releases only a small fraction of the energy available in the complete oxidation of glucose. In aerobic organisms, more than 90% of the ATP made during glucose catabolism results from the oxidation of pyruvate.

B. Oxidation of Pyruvate: The Citric Acid Cycle

In higher organisms, the oxidation of pyruvate takes place in subcellular, membranous organelles known as mitochondria. Because mitochondria are responsible for the synthesis of most of the ATP in nonphotosynthetic tissue, they are often referred to as the powerhouses of cells. Mitochondrial ATP synthesis is called oxidative phosphorylation since it is linked indirectly to oxidative reactions. In the complete oxidation of pyruvate, there are five oxidation–reduction reactions. Three of these reactions are oxidative decarboxylations. The electron acceptor (oxidizing agent) for four of the reactions is NAD^+; the oxidizing agent for the fifth is flavin adenine dinucleotide, or FAD.

Knowing the oxidation–reduction potentials of the reactants in an oxidation–reduction reaction permits the ready calculation of the standard free energy change for the reaction. It may be shown that

$$\Delta G_0' = -nF\Delta E_0', \qquad (1)$$

where n is the number of electrons transferred in the reaction, F is Faraday's constant (23,060 cal/V-equivalent), and $\Delta E_0'$ is the difference between the E_0' value of the oxidizing agent and that of the reducing agent.

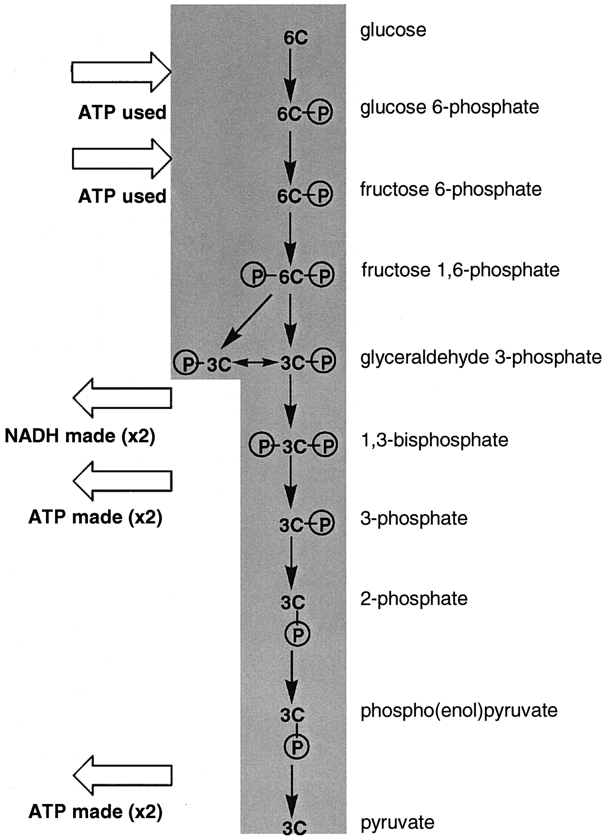


FIGURE 4 A view of glycolysis. Glucose, a six-carbon sugar, is cleaved and oxidized to two molecules of pyruvate. There is the net synthesis of two ATP per glucose oxidized and two NADH are also formed.

The reduced form of NAD^+, NADH, is a strong reducing agent. The E_0' at pH 7.0 of the NAD^+–NADH couple is −340 mV, which is equivalent to that of molecular hydrogen. E_0 is the potential when the concentrations of the oxidized and reduced species of an oxidation–reduction pair are equal. Reduced FAD, $FADH_2$, is a weaker reductant than NADH, with an E_0' (pH 7.0) of about 0 V. In contrast, molecular oxygen is a potent oxidizing agent and fully reduced oxygen, water, is a very poor reducing agent. The E_0' (pH 7.0) for the oxygen–water couple is +815 mV.

The oxidation of NADH and $FADH_2$ results in the reduction of oxygen to water:

$$H^+ + NADH + \tfrac{1}{2}O_2 \rightarrow NAD^+ + H_2O \qquad (2)$$

and

$$FADH_2 + \tfrac{1}{2}O_2 \rightarrow FAD + H_2O. \qquad (3)$$

In both cases two electrons are transferred to oxygen, so that the n in Eq. (1) is equal to 2. Under standard conditions, the oxidation of 1 mol of NADH by oxygen liberates close to 53 kcal, whereas the $\Delta G_0'$ for that of $FADH_2$ is −38 kcal/mol. These two strongly exergonic reactions provide the energy for the endergonic synthesis of ATP.

The details of carbon metabolism in the citric acid cycle are beyond the scope of this article. In brief, pyruvate is first oxidatively decarboxylated to yield CO_2, NADH, and an acetyl group attached in an ester linkage to a thiol on a large molecule, known as coenzyme A, or CoA. (See Fig. 2.) Acetyl CoA condenses with a four-carbon dicarboxylic acid to form the tricarboxylic acid citrate. Free CoA is also a product (Fig. 6). A total of four oxidation–reduction reactions, two of which are oxidative decarboxylations, take place, which results in the generation of the three remaining NADH molecules and one molecule of $FADH_2$. The citric acid cycle is a true cycle. For each two-carbon acetyl moiety oxidized in the cycle, two CO_2 molecules are produced and the four-carbon dicarboxylic acid with which acetyl CoA condenses is regenerated.

The mitochondrial inner membrane (Fig. 7) contains proteins that act in concert to catalyze NADH and $FADH_2$ oxidation by molecular oxygen. [See reactions (2) and (3) above.] These reactions are carried out in many small steps by proteins that are integral to the membrane and that undergo oxidation–reduction. These proteins make up what is called the mitochondrial electron transport chain. Components of the chain include iron proteins (cytochromes and iron–sulfur proteins), flavoproteins (proteins that contain flavin), copper, and quinone binding proteins.

The oxidation of NADH and $FADH_2$ by molecular oxygen is coupled in mitochondria to the endergonic synthesis of ATP from ADP and P_i. For many years the nature of the common intermediate between electron transport and ATP synthesis was elusive. Peter Mitchell, who received a Nobel Prize in chemistry in 1978 for his extraordinary insights, suggested that this common intermediate was the proton electrochemical potential. He proposed in the early

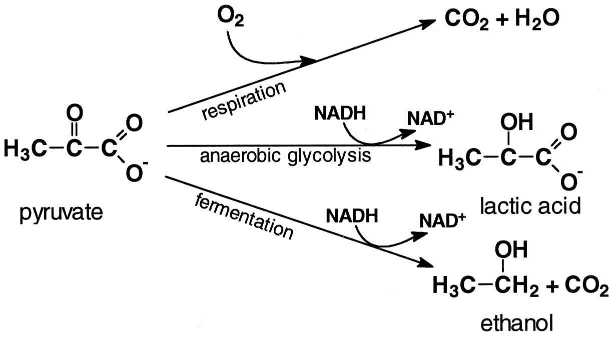


FIGURE 5 Fates of pyruvate. In yeasts under anaerobic conditions, pyruvate is decarboxylated and reduced by the NADH formed by glycolysis to ethanol. In anaerobic muscle, the NADH generated by glycolysis reduces pyruvate to lactic acid. When O_2 is present, pyruvate is completely oxidized to CO_2 and water.

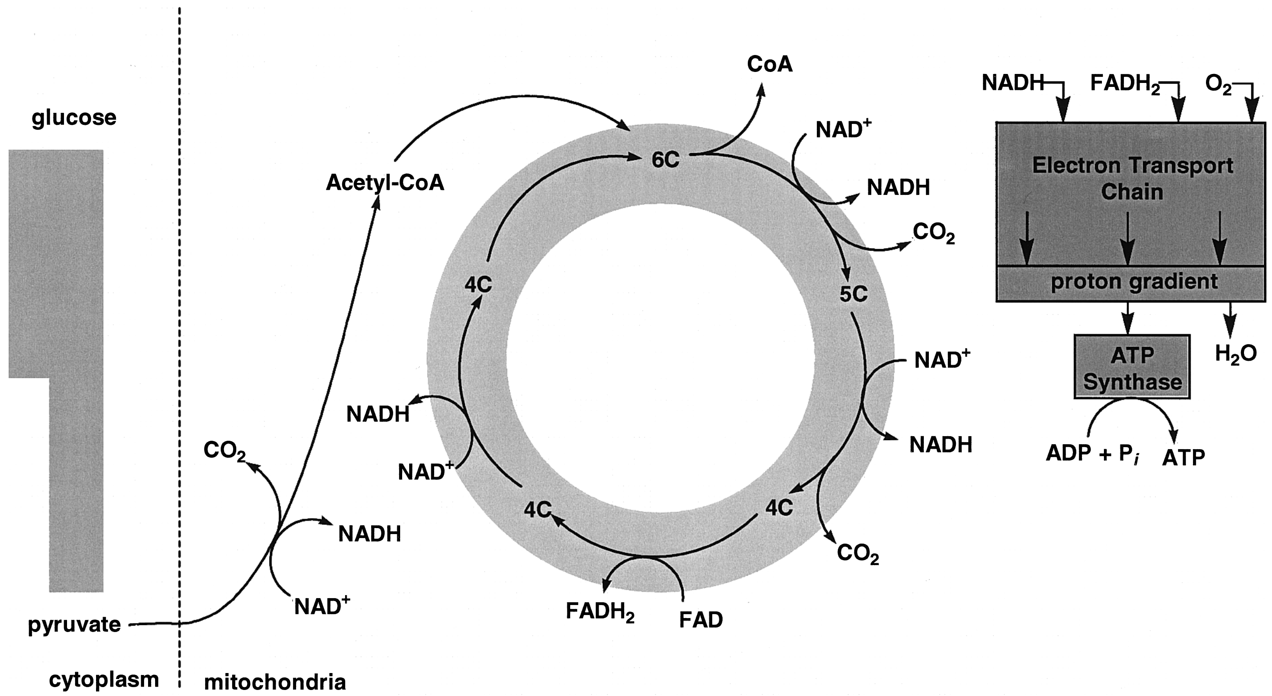


FIGURE 6 A view of the oxidation of pyruvate. The oxidation of pyruvate generates three CO_2, four NADH, and one $FADH_2$. The oxidation of NADH and $FADH_2$ by the mitochondrial electron transport chain is exergonic and provides most of the energy for ATP synthesis.

1960s that electron transport through the mitochondrial chain is obligatorily linked to the movement of protons across the inner membrane of the mitochondrion. In this way, part of the energy liberated by oxidative electron transfer is conserved in the form of the proton electrochemical potential. This potential, $\Delta\mu_{H^+}$, is the sum of contributions from the activity gradient and that of the electrical gradient:

$$\Delta\mu_{H^+} = RT \ln\left([H^+]_a/[H^+]_b\right) + F\Delta\varphi, \qquad (4)$$

where R is the gas constant; T, the absolute temperature; a and b, the aqueous spaces bounded by the membrane; F, Faraday's constant; and $\Delta\varphi$, the membrane potential. As Mitchell suggested, the mitochondrial inner membrane is poorly permeated by charged molecules, including protons. The membrane thus provides an insulating layer between the two aqueous phases it separates. Thus the transport of protons across the membrane generates an electrochemical potential. In the case of mitochondria, the membrane potential is the predominant component of the electrochemical of the proton. The total $\Delta\mu_{H^+}$ in actively respiring mitochondria is on the order of −200 mV, if one uses the convention that the inside space bounded by the membrane is negative.

Electron transport from NADH and $FADH_2$ to oxygen provides the energy for the generation of the electrochemical potential of the proton. The flow of protons down this potential is exergonic and is the immediate source of energy for ATP synthesis. The proton-linked synthesis of ATP is catalyzed by a complex enzyme called ATP synthase. Remarkably similar enzymes are located in the coupling membranes of bacteria, mitochondria, and chloroplasts, the intracellular sites of photosynthesis in higher plants. Even though the reaction that they catalyze seems relatively straightforward (see Fig. 2), the ATP synthases contain a minimum of 8 different proteins and a total of about 20 polypeptide chains.

ATP is formed in the aqueous space bounded by the mitochondrial inner membrane. This space is known as the matrix (see Fig. 7). Most of the ATP generated within mitochondria is exported to the cytoplasm where it is used to drive energy-dependent reactions. The ADP and P_i formed in the cytoplasm must then be taken up by the mitochondria. The inner membrane contains specific proteins that mediate the export of ATP and the import of ADP and P_i. One transporter catalyzes counterexchange transport of ATP out of the matrix with ADP in the cytoplasm into the matrix (Fig. 8). At physiological pH, ATP bears four negative charges, and ADP, three. Thus, the one-to-one exchange transport of ATP with ADP creates a membrane potential that is opposite in sign of that created by electron-transport-driven proton translocation. ATP/ADP transport costs energy and the direction of transport is poised by the proton membrane potential. In addition, phosphate

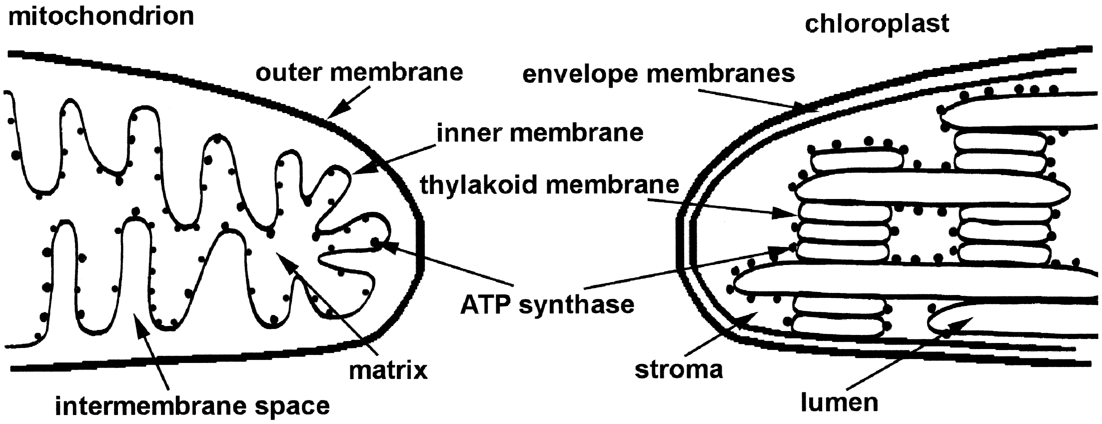


FIGURE 7 Diagrams of the structures of mitochondria and chloroplasts. The inner membrane of mitochondria and the thylakoid membrane of chloroplasts contain the electron transport chains and ATP synthases. Note that the orientation of the inner membrane is opposite that of the thylakoid membrane.

uptake into mitochondria is coupled to the electrochemical proton potential. The phosphate translocator (see Fig. 8) catalyzes the counterexchange transport of $H_2PO_4^{2-}$ and hydroxide anion (OH^-). The outward movement of OH^- causes acidification of the matrix, whereas the direction of proton transport driven by electron transport is out of the mitochondrial matrix and results in an increase in the pH of the matrix.

In the total oxidation of glucose to CO_2 and water, six CO_2 are released and six O_2 are reduced to water. For each pyruvate oxidized, four NADH and one $FADH_2$ are generated. Since two molecules of pyruvate are derived by means of glycolysis from one molecule of glucose, a total of eight NADH and two $FADH_2$ are formed by pyruvate oxidation. Four electrons are required for the reduction of O_2 to two molecules of H_2O. Thus, pyruvate oxidation accounts for the reduction of five of the six molecules of O_2 in the complete oxidation of glucose. The sixth O_2 is reduced to water by electrons from the NADH formed by the oxidation of triose phosphate in glycolysis.

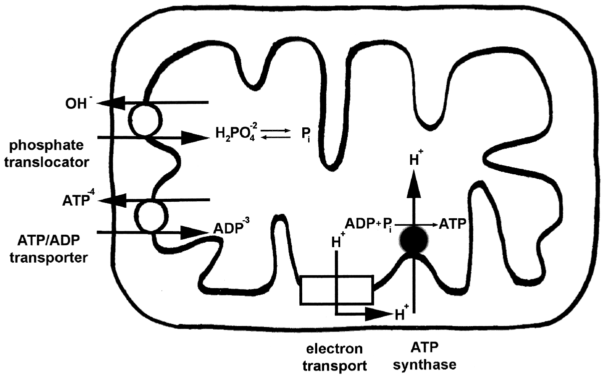


FIGURE 8 ATP, ADP, and P_i transport in mitochondria. ATP is formed inside mitochondria. Most of the ATP is exported to the cytoplasm where it is cleaved to ADP and P_i. The mitochondrial inner membrane contains specific proteins that mediate not only ATP release coupled to ADP uptake, but also P_i uptake linked to hydroxide ion (OH^-) release.

Fermentation, or anaerobic glycolysis, yields but 2 mol of ATP per 1 mol of glucose catabolized. In contrast, complete oxidation of glucose to CO_2 and water yields about 15 times more ATP. Thus, it is understandable why yeasts and some bacteria consume more glucose under anaerobic conditions than when oxygen is present.

In animals, glucose is normally completely oxidized. During strenuous exercise, however, the demand for oxygen by muscle tissues can outstrip its supply and the tissue may become anaerobic. Muscle contraction requires ATP, and rapid breakdown of glucose and its storage polymer, glycogen, takes place under anaerobiosis. Glycolysis would stop quickly if the NADH produced by the oxidation of triose phosphate were not converted back to NAD^+. In muscle cells under O_2-limited conditions, pyruvate is reduced by NADH to lactic acid (see Fig. 5), a source of muscle cramps during exercise. At rest, lactic acid is converted back to glucose in the liver and kidneys and returned to muscle tissues where it stored in the form of glycogen.

C. Oxidation of Fats and Oils, Major Metabolic Fuels

Fats and oils are ubiquitous biological molecules that are major energy reserves in animals and developing plants. Fats and oils are esters of glycerol, a three-carbon compound with hydroxyl groups on all three carbons, and carboxylic acids with long hydrocarbon chains. The most common fats and oils contain fatty acids with straight chains with an even number of carbon atoms. Most often, the total number of carbons in a fatty acid in a triglyceride ranges from 14 to 18. The difference between a fat and an

oil is simply melting temperature. Oils are liquid at room temperature, whereas fats are solid. Familiar examples are olive oil and butter.

The most significant reason for this difference in melting temperatures between fats and oils is the degree of unsaturation (double bonds) of the fatty acids they contain. The introduction of double bonds into a hydrocarbon chain causes perturbations in the structure of the chain that decrease its ability to pack the chains closely into a solid structure. Olive oil contains far more unsaturated fatty acids than butter does and is thus a liquid at room temperature and even in the cold.

Regardless of the physical properties of triglycerides, they are the long-term energy reserves of higher organisms. Consider the fact that the complete oxidation of triglycerides to CO_2 and water yields 9 kcal/g, whereas that of the carbohydrate storage polymers, starch and glycogen, yields just 4 kcal/g. When it is also remembered that fats and oils shun water, but glycogen and starch are more hydrophilic, triglycerides have an additional advantage over the glucose polymers as deposits of potential free energy. As hydrophobic moieties, fats and oils require less intracellular space than that required by the glucose polymers.

The first step in the breakdown of triglycerides (Fig. 9) is their conversion by hydrolysis to their components, glycerol and fatty acids. Glycerol is a close relative of the three-carbon compounds involved in the catabolism of glucose and may be completely oxidized to CO_2 and water by glycolysis and the tricarboxylic acid cycle.

The fatty acids are first converted to CoA derivatives at the expense of the hydrolysis of ATP and then transported into mitochondria where they are broken down sequentially, two carbon units at a time, by a pathway known as β-oxidation (see Fig. 9). The fatty acyl CoA derivatives undergo oxidation at the carbon that is β to the carboxyl carbon from that of a saturated carbon–carbon bond to that of an oxo-saturated carbon bond. Enzymes that contain FAD or use NAD^+ as the electron acceptors catalyze these reactions. As is the case in the oxidation of carbohydrates, the NADH and $FADH_2$ generated by the β-oxidation of fatty acids are converted to their oxidized forms by the mitochondrial electron transport chain, which results in the formation of ATP by oxidative phosphorylation.

Once β-oxidation is complete, the terminal two carbons of the fatty acid chain are then released as acetyl CoA. Oxidation and cleavage of the fatty acid continue until it is entirely converted to acetyl CoA. The conversion of a saturated fatty acid with 18 carbon atoms to 9 acetyl CoA produces 8 NADH and 8 $FADH_2$. The acetyl CoA is burned by the citric acid cycle to generate more ATP. The high caloric content of fats pays off to cells in the yield of ATP.

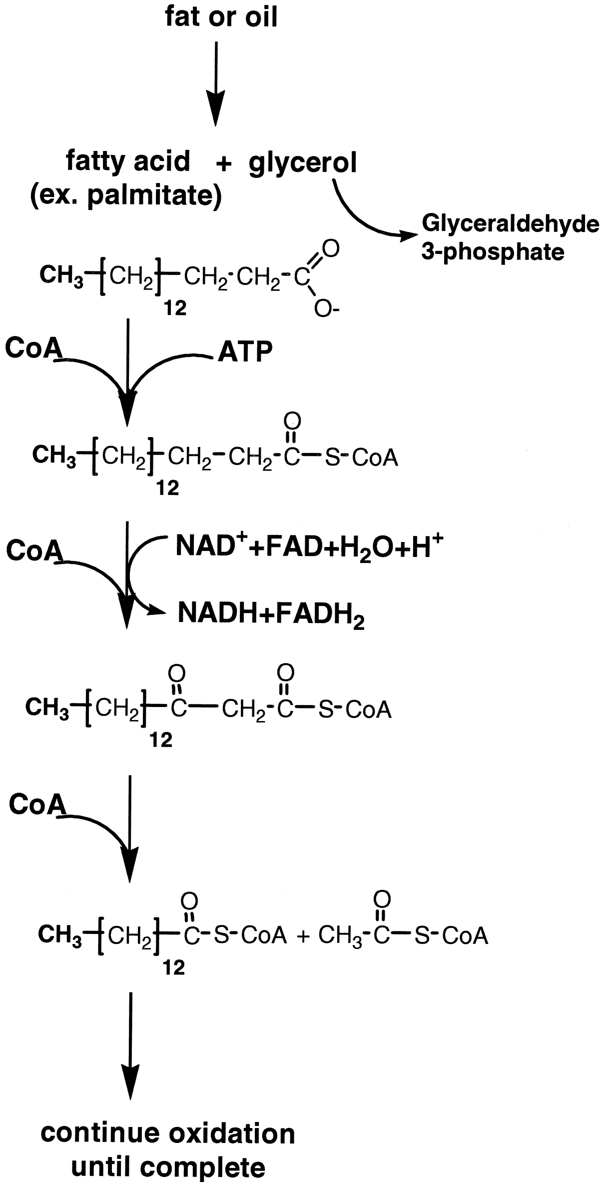


FIGURE 9 Oxidation of fatty acids. Fats and oils are hydrolyzed to form glycerol and fatty acids. CoA derivatives of the fatty acids are oxidized in mitochondria by NAD^+ and FAD to β-oxo-derivatives. CoA cleaves these derivatives to yield acetyl CoA and a fatty acid CoA molecule that is two carbons shorter. The process continues until the fatty acid has been completely converted to acetyl CoA. The acetyl moiety is oxidized in the citric acid cycle to CO_2 and water. The complete oxidation of a fatty acid of about the same molecular weight of glucose yields four times more ATP than that of glucose.

D. Catabolism of Proteins and Amino Acids

In addition to containing carbohydrates and fats, diets may be rich in proteins. The catabolism of proteins results in the generation of their component parts, amino acids. When the dietary amino acid requirements of an individual are

satisfied, the excess amino acids in the diet are catabolized to CO_2 and water as a source of energy. Some amino acids are degraded to molecules that feed directly into glycolysis, and others result in the production of acetyl CoA. Excess nitrogen resulting from the catabolism of amino acids and other compounds that contain nitrogen is excreted in mammals in urine in the form of the simple organic compound urea. Some amino acids are precursors in the biosynthesis of other organic molecules.

E. Summary

In summary, organisms such as humans and other animals, many bacteria, fungi, and nongreen plants derive the energy they must have to power life from foodstuffs they obtain from their environment. The degradation of carbohydrates and the oxidation of fats are the major sources of energy for heterotrophic (other feeding) organisms. How these molecules that are essential to life are generated is the next subject considered.

II. PHOTOSYNTHESIS

From a purely thermodynamic standpoint, life is an improbable event. Consider, for example, the complex structures of organisms, not only at the macroscopic level, but also at the microscopic and atomic levels. These ordered structures can be formed and maintained only by the expenditure of energy. Within the ecosystem that we call the earth, the organic nutrients necessary to sustain the life of heterotrophs such as us are provided directly and indirectly by photosynthesis.

In both quantitative and qualitative terms photosynthesis is the most significant biological process on Earth. Approximately 2×10^{11} tons of carbon dioxide are converted to organic compounds each year. It is to photosynthesis in prehistoric times that we owe the reserves of fossil fuels. The oxygen that we breathe is a direct result of photosynthesis, now and in prehistory.

If the earth were an isolated system in a thermodynamic sense, life would be in jeopardy in that the energy reserves for life would be consumed. Without the input of energy from a source external to the earth, the planet must tend toward achieving equilibrium within its environment.

Fortunately, the earth is not an isolated system. The hydrogen fusion reactor of the Sun bathes our planet in electromagnetic radiation, including visible light. A fraction of the solar energy that impinges on Earth is converted by photosynthesis to chemical energy in the form of organic molecules that heterotrophic organisms use to satisfy their continued need for energy. The process by which light energy is used to drive the otherwise unfavorable synthesis of these organic molecules is called photosynthesis.

Although some bacteria carry out photosynthesis without the evolution of oxygen, this article deals solely with oxygenic photosynthesis that takes place in higher plants and algae. In a purely formal sense, oxygenic photosynthesis may be represented as the reverse of the oxidative breakdown of a six-carbon carbohydrate, such as glucose. An equation that describes photosynthesis in part illustrates this relationship:

$$6CO_2 + 12H_2O \rightarrow C_6H_{12}O_6 + 6O_2 + 6H_2O, \qquad (5)$$

where $C_6H_{12}O_6$ refers to a six-carbon sugar. This equation in reverse describes the oxidative catabolism of a six-carbon sugar such as glucose. Under standard conditions, the complete oxidation of glucose liberates 686 kcal/mol; the synthesis of a mole of glucose from carbon dioxide and water thus minimally requires the input of an equivalent amount of energy. In photosynthesis, visible light provides this energy. When it is considered that the only source of carbon for the tens of thousands of organic compounds synthesized in green plants is from the assimilation of carbon dioxide by means of photosynthesis, the inadequacy of Eq. (5) to describe photosynthesis, despite its usefulness, is readily apparent.

Inspection of Eq. (5) reveals that photosynthesis is an oxidation–reduction process. Simply put, photosynthesis is the light-driven reduction of carbon dioxide to the oxidation–reduction level of a carbohydrate by using water as the electron and hydrogen donor. In the process, water is oxidized to molecular oxygen. As stated previously, water is a very poor reducing agent. However, water at an effective concentration of 55 *M* is readily available in the biosphere. Although organic compounds and inorganic molecules such as hydrogen sulfide are more powerful reducing agents than water is, their use in photosynthesis as the source of electrons for photosynthesis is restricted to certain species of bacteria. The thermodynamically very unfavorable reduction of carbon dioxide by water is driven by light.

A. Light Reactions

How the electromagnetic energy of light is converted to chemical energy in the form of reduced organic molecules is complex. Nonetheless, the first principles of energy conservation and conversions in photosynthesis may be simply depicted. All higher photosynthetic organisms contain two forms of the green pigment chlorophyll. More than 99% of the chlorophyll in chloroplasts, the organelles in which photosynthesis takes place, functions in a passive, purely physical manner. Organized in specific pigment–protein complexes within the photosynthetic membrane, these chlorophylls absorb visible light and transfer excitation energy to nearby chlorophylls with efficiencies very close to 100%. In a real sense, more than 99% of the

chlorophylls function only to gather light and as such they are often referred to as light-harvesting chlorophylls.

Within picoseconds of the harvesting, the excitation energy is transferred to specialized chlorophyll molecules called reaction center chlorophylls. These reaction center chlorophylls are identical to the majority of the light-harvesting chlorophylls. Yet, rather than acting in a passive manner when they are excited, the reaction center chlorophylls perform photochemistry. The two reaction center chlorophylls are termed P700 and P680. The "P" stands for pigment and the numbers refer to their absorption maxima, in nanometers, in the red region of the spectrum. The reaction center chlorophylls were first detected by light-induced bleaching at 680 and 700 nm. When the reaction center chlorophylls are excited, either directly or by resonance energy transfer from excited light-harvesting chlorophylls, an electron is transferred from the reaction center chlorophyll ensemble to an electron acceptor. These light-driven oxidation–reduction reactions occur within picoseconds and can operate with a quantum efficiency that is close to 100%. The reactions may be written as follows:

$$\text{P700}^* + \text{FeS} \rightarrow \text{P700}^+ + \text{FeS}^- \tag{6}$$

and

$$\text{P680}^* + \text{Q} \rightarrow \text{P680}^+ + \text{Q}^-, \tag{7}$$

where the asterisks indicate the first excited singlet state of the reaction center chlorophyll, and FeS and Q are the redox active part of an iron–sulfur protein and a quinone, respectively, the first stable electron acceptors. $P700^+$ and $P680^+$ are chlorophyll cation radicals and Q^- is a half reduced quinone and FeS^- is a reduced iron-sulfur protein. The reactions shown in Eqs. (6) and (7) cannot take place, in the direction shown, in the dark when the reaction center chlorophylls are in the unexcited, ground state. The $\Delta G_0'$ for both these reactions is approximately +24 kcal/mol. The excited reaction center chlorophylls are, however, much stronger reducing agents than the ground state chlorophylls are. The E_0' of P700* is about 1.3 V more reducing than that of P700 in the ground state. These two electron transfer reactions are the only light-driven reactions in photosynthesis and they set the entire process in motion. The electron transport chain of chloroplasts is illustrated in Fig. 10.

Specific light-harvesting chlorophyll–protein complexes are associated with the reaction center chlorophyll–protein complexes in assemblies known as photosystems. Photosystem I (PS I) contains P700 and the FeS acceptor, and photosystem II (PS II), P680 and the quinone acceptor. Electron transfer in PS I generates a relatively weak oxidizing agent ($P700^+$, $E_0' = +430$ mV) and a strong reductant (FeS^-, $E_0' = -600$ mV). The primary reductant generated in photosynthesis is nicotinamide adenine dinucleotide phosphate ($NADP^+$), which, as the name suggests, differs from NAD^+ by a single phosphate. While the physical properties of $NADP^+$ and NAD^+ are very similar, enzymes that use these pyridine nucleotides as substrates can discriminate between them by at least a factor of 1000. In general NAD^+ is used in catabolic metabolism as we have seen for glycolysis and the tricarboxylic acid cycle. The reduced form of $NADP^+$, NADPH, is, in contrast,

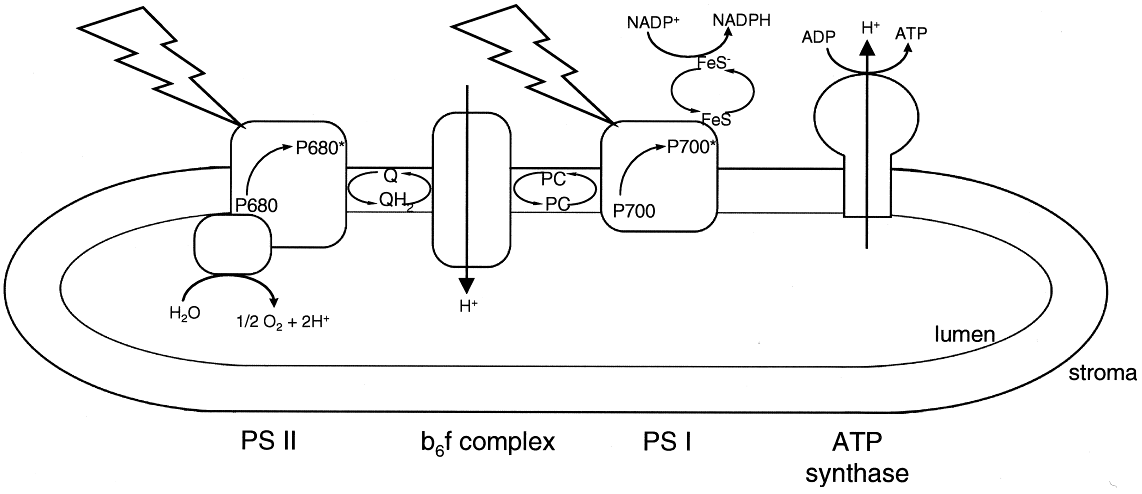


FIGURE 10 Electron transport and ATP synthesis in chloroplasts. The jagged arrows represent light striking the two photosystems (PS I and PS II) in the thylakoid membrane. Other members of the electron transport chain shown are a quinone (Q), the cytochrome complex (b_6f), plastocyanin (PC), and an iron–sulfur protein (FeS). The chloroplast ATP synthase is shown making ATP at the expense of the electrochemical proton gradient generated by electron transport.

used in biosynthesis, or anabolic metabolism. The E_0' of the $NADP^+$–NADPH redox pair is −340 mV. Thus, electron transfer from the reduced iron–sulfur protein of PS I to $NADP^+$ is energetically a very favorable spontaneous reaction. It is NADPH that provides the electrons for CO_2 reduction. The ultimate electron donor is water.

Two water molecules are oxidized by PS II to yield four protons and molecular oxygen. Water is a very weak reducing agent. Thus, a strong oxidizing agent is needed for water oxidation. $P680^+$ fits the bill. The midpoint potential of the $P680^+$–P680 redox pair is on the order of +1 V. Since the water–oxygen redox couple has an E_0' of +0.815 V, the oxidation of water by $P680^+$ is an energetically spontaneous reaction. Water oxidation is catalyzed by a manganese-containing enzyme that is plugged into the energy-converting thylakoid membrane.

So far, we have seen that the reduced FeS protein of PS I is converted to its oxidized form by passing electrons eventually to $NADP^+$. In PS II, $P680^+$ is reduced to P680 with electrons extracted from water. For electron transport to continue, the electron acceptor of PS II, Q^-, and the electron donor of PS I, $P700^+$, must be oxidized and reduced, respectively. The redox potential of the Q–Q^- couple is about +0.05 V, whereas that of $P700^+$–P700 is near +0.450 V. Thus, electron transport from Q^- to $P700^+$ is energetically spontaneous with a free energy of 9.3 kcal/mol for each electron transferred.

Electron transport from Q^- to $P700^+$ is mediated by a quinone, iron–sulfur, and a cytochrome protein complex in the thylakoid membrane. This protein, the cytochrome $b_6 f$ complex, is remarkably similar to the cytochrome bc_1 complex of the mitochondrial electron transport chain.

B. CO_2 Reduction

Linear electron transport in oxygenic photosynthesis is the reduction of $NADP^+$ to NADPH by water, which results in the formation of molecular oxygen:

$$2H_2O + 2NADP^+ \rightarrow O_2 + NADPH + 2H^+. \quad (8)$$

NADPH is incapable of reducing CO_2 by itself; ATP is also required. The CO_2 acceptor in photosynthesis is the five-carbon, phosphorylated sugar ribulose 1,5-bisphosphate. CO_2 cleaves this sugar into 2 mol of the three-carbon sugar acid 3-phosphoglycerate, a compound that is also an intermediate in glycolysis. The enzyme that catalyzes this reaction, ribulose 1,5-bisphosphate carboxylase/oxygenase, or rubisco, is present in very high concentrations within chloroplasts, which makes it among the most abundant proteins in the biosphere.

Recall that in glycolysis one of the two steps in which ATP is formed is the conversion of 1,3-bisphosphoglycerate to 3-phosphoglycerate. The acyl phosphate at the 1-position of the bisphosphorylated sugar acid is transferred to ADP to form ATP. The conversion of 3-phosphoglycerate to carbohydrates occurs by a pathway that is essentially the reverse of glycolysis. It must be emphasized, however, that glycolysis and photosynthetic carbon metabolism take place in separate intracellular compartments. Glycolysis occurs in the cytoplasm and uses NAD^+ as the electron acceptor. The photosynthetic reduction of 3-phosphoglycerate occurs inside chloroplasts in the aqueous space known as the stroma. The enzymes in the two compartments are not the same even though they catalyze similar reactions. For example, the triose phosphate dehydrogenase in the cytoplasm is very specific for NAD^+, whereas that in the chloroplast stroma is equally specific for NADPH.

Therefore, ATP is required for the reduction by NADPH of 3-phosphoglycerate to the oxidation level of a carbohydrate:

$$\text{ATP} + \text{3-phosphoglycerate} \rightarrow \text{ADP} + \text{1,3-bisphosphoglycerate}, \quad (9)$$

and the bisphosphoglycerate is in turn reduced by NADPH:

$$\text{NADPH} + \text{H}^+ + \text{1,3-bisphosphoglycerate} \rightarrow \text{NADP}^+ + \text{P}_\text{i} + \text{glyceraldehyde 3-phosphate}. \quad (10)$$

Since two 3-phosphoglycerates are generated for each CO_2 assimilated, two NADPH and two ATP are required for reduction. This reaction is the only one in photosynthetic carbohydrate metabolism that is an oxidation–reduction reaction.

Glyceraldehyde 3-phosphate is a sugar phosphate and may be readily converted within chloroplasts to many sugars and the glucose polymer starch. Some of the glyceraldehyde 3-phosphate is used in a complex series of reactions to regenerate the five-carbon acceptor of CO_2, ribulose 1,5-bisphosphate. In the process, one phosphate is cleaved from one of the sugar phosphate intermediates. Thus, ribulose 5-phosphate, the product of the cycle, must be phosphorylated by using ATP as the phosphoryl donor. As a consequence, three ATP and two NADPH are required for each CO_2 taken up.

Photosynthesis must satisfy the energy requirements of all living tissues in plants, including roots, stems, and developing fruit. Up to 75% of the triose phosphate formed is exported from the chloroplasts in leaf cells to the cytoplasm where it is converted to sucrose, a major product of photosynthesis. In most plants, sucrose is transported to the rest of the plant where it is either stored as starch or broken down by glycolysis and the citric acid cycle in exactly the same way as it is in animals to produce the ATP needed to sustain life.

C. ATP Synthesis

ATP synthesis in chloroplasts is called photophosphorylation and is similar to oxidative phosphorylation in mitochondria. The light-driven transport of electrons from water to $NADP^+$ is coupled to the translocation of protons from the stroma across the thylakoid membrane (the green, energy-converting membrane) into the lumen. Electron transport from Q^- to $P700^+$ is exergonic. Part of the energy released by electron transport is conserved by the formation of an electrochemical proton gradient. The cytochrome b_6f complex of chloroplasts functions not only in electron transport, but also in proton translocation.

The active site of the oxygen-evolving enzyme is arranged so that the protons formed during water oxidation are released into the thylakoid lumen. These protons contribute to the electrochemical proton potential. The thylakoid membrane contains a protein that functions to transport Cl^- across the membrane. Proton accumulation in the thylakoid lumen is electrically balanced in large part by Cl^- uptake. As a result, thylakoids accumulate HCl and the membrane potential across the membrane is low. The pH inside the lumen during steady-state photosynthesis is about 5.0.

One of the earliest experiments that supported the hypothesis that ATP synthesis and electron transport were linked by the electrochemical proton potential was carried out with isolated thylakoid membranes. Thylakoid membranes were placed in a buffer at pH 4.0 and after a few seconds the pH was rapidly increased to 8.0, which resulted in the formation of a proton activity gradient. This artificially formed gradient was shown to drive the synthesis of ATP from ADP and P_i. The experiments were carried out in the dark so that the possibility that electron transport contributed to the ATP synthesis was excluded. Thus, a proton activity gradient was proven capable of driving ATP synthesis.

The thylakoid membrane enzyme that couples ATP synthesis to the flow of protons down their electrochemical gradient is called the chloroplast ATP synthase (see Fig. 10). This enzyme has remarkable similarities to ATP synthases in mitochondria and certain bacteria. For example, the β subunits of the chloroplast ATP synthase have 76% amino acid sequence identity with the β subunits of the ATP synthase of the bacterium *E. coli*.

The reaction catalyzed by ATP synthases is

$$n\mathrm{H}_\mathrm{a}^+ + \mathrm{ADP} + \mathrm{P_i} + \mathrm{H}^+ \rightarrow n\mathrm{H}_\mathrm{b}^+ + \mathrm{ATP} + \mathrm{H_2O}, \quad (11)$$

where n is the number of protons translocated per ATP synthesized, probably three or four, and a and b refer to the opposite sides of the coupling membrane. Provided the electrochemical proton potential is high, the reaction is poised in the direction of ATP synthesis. In principle, when the proton potential is low, ATP synthases should hydrolyze ATP and cause the pumping of protons across the membrane in the direction opposite that which occurs during ATP synthesis. ATP-dependent proton transport by the ATP synthase is of physiological significance in *E. coli* under anaerobic conditions in that it generates the electrochemical proton potential across the plasma membrane of the bacterium. This potential is used for the active uptake of some carbohydrates and amino acids.

In contrast, ATP hydrolysis by the chloroplast ATP synthase in the dark has no physiological role and would be wasteful. In fact, the rate of ATP hydrolysis by the ATP synthase in thylakoids in the dark is less than 1% of the rate of ATP synthesis in the light. Remarkably, within 10–20 msec after the initiation of illumination, ATP synthesis reaches its steady-state rate. Thus, the activity of the chloroplast ATP synthase is switched on in the light and off in the dark. In addition to being the driving force for ATP synthesis, the electrochemical proton potential is involved in switching the enzyme on. Structural perturbations of the enzyme induced by the proton potential overcome inhibitory interactions with bound ADP as well as with a polypeptide subunit of the synthase. An additional regulatory mechanism that is unique to the chloroplast ATP synthase is reductive activation. Reduction of a disulfide bond in a subunit of the chloroplast ATP synthase to a dithiol enhances the rate of ATP synthesis, especially at physiological values of the proton potential. The electrons for this reduction are derived from the chloroplast electron transport chain.

III. ORIGIN OF MITOCHONDRIA AND CHLOROPLASTS

In animal, yeast, and fungal cells, DNA is present in two organelles, the nucleus and the mitochondria. In plant and algal cells, DNA is present in plastids (of which chloroplasts are one example) as well as in mitochondria and the nucleus. Unlike the DNA in the nucleus, which is packaged into chromosomes, plastid DNA and mitochondrial DNA are circular and thus resemble the DNA in prokaryotes (e.g., bacteria).

Mitochondrial DNA is small and codes for relatively few mitochondrial proteins. Although mitochondria contain their own protein synthesis machinery, the majority of the hundreds of mitochondrial proteins are coded for by nuclear genes. These proteins are synthesized in the cytoplasm and imported into the mitochondria. Plastid DNA is somewhat larger than that of the mitochondrion and contains the genetic information for more chloroplast proteins. However, as is the case for mitochondria, most of the proteins in a chloroplast are coded by nuclear genes

and are synthesized in the cytoplasm. Proteins destined for mitochondria and chloroplasts have an extension on their *N*-terminal end that targets the proteins to the correct organelle and to the correct place within the organelle. These extensions, which, like the remainder of the proteins, are composed of amino acids, are usually cleaved off as the proteins find their proper place within the organelle. Remarkably, some proteins composed of more than one polypeptide may contain a polypeptide coded for by nuclear DNA and synthesized in the cytoplasm and another polypeptide that is coded for by mitochondrial or chloroplast DNA. Ribulose 1,5-bisphosphate carboxylase/oxygenase is a prominent example of such a protein in chloroplasts.

The discovery that mitochondria and chloroplasts contain DNA, coupled with a wealth of sequence information about both DNA and proteins, added credence to the notion that these organelles arose from the engulfment of unicellular organisms by a primitive nucleated cell. Mitochondria may have been derived from a bacterium, and chloroplasts, from a unicellular alga. After the engulfment events, genes in the bacterium and alga coding for proteins that duplicated those in the nuclear genomes of the hosts were lost and other genes were transferred from the bacterial and algal genomes to the genomes of the hosts.

The distribution of proteins and lipids within biological membranes is asymmetric. Thus, one side of a membrane is distinct from the other. The coupling membranes of mitochondria and chloroplasts are opposite to each other. Protons are ejected from mitochondria during respiratory electron transport but are taken up by thylakoids during light-driven electron transport. The catalytic portion of the ATP synthase is located on the outside of the thylakoid membranes, whereas that of the mitochondrial ATP synthase is present on the inside of the inner membrane. As seen in Fig. 7, the orientation of the coupling membranes of mitochondria and chloroplasts is consistent with the hypothesis that these organelles are of bacterial and algal origin.

Each membrane in a cell has its distinct set of proteins and lipids. The most common membrane lipids are phospholipids. Phospholipids are diglycerides. Two of the three hydroxyls of glycerol are linked to long-chain fatty acids by ester bonds. The third position is occupied by phosphate. A number of different polar substituents are linked to the phosphate by anhydride bonds. The phospholipid composition of the mitochondrial inner membrane is virtually the same in plant mitochondria as in animal mitochondria and resembles that in the plasma membrane of some bacteria. The lipids in chloroplast membranes are very distinctive. The phospholipid content is unusually low and about 80% of the membrane lipids in thylakoids are diglycerides that have one or two galactose (a six-carbon sugar) on the third position of the glycerol. Galactosyldiglycerides are absent in the membranes of animal, yeasts, and fungi but are present in the photosynthetic membranes of all organisms that carry out oxygenic photosynthesis. The lipid compositions of mitochondrial and chloroplast membranes are consistent with the engulfment hypothesis for the origin of these organelles.

IV. ILLUSTRATIONS OF THE USES OF ATP: ION TRANSPORT, BIOSYNTHESIS, AND MOTILITY

ATP powers most of the endergonic processes in cells. How the potential energy of the phosphoanhydride bond of ATP may be used to drive otherwise unfavorable reactions (Fig. 11) is discussed in this section. This discussion focuses on three major uses of ATP: the generation of ion gradients, biosynthesis, and movement.

A. Ion Transport

The plasma membrane is the barrier that separates the cytoplasm of cells from the exterior medium. All cells maintain a membrane potential that is negative. There is an excess of positive charge in the external medium in comparison with that in the cytoplasm. The membrane potential in plant cells can be as high as −200 mV. Energy is required to generate and maintain the membrane potential.

All cells maintain gradients in ions across the plasma membrane. The intracellular K^+ concentration is higher than that of the extracellular medium, and the concentration of Na^+, much lower. The free Ca^{2+} concentration in the cytoplasm is maintained at very low levels, 1000-fold or more below the extracellular Ca^{2+} concentration. Often the intracellular proton concentration can be quite different from that in the medium. The pH in the cytoplasm of plant cells is close to 7.0, whereas that in the medium is about 5.0. Energy is needed to generate and maintain these ionic disequilibria. For example, the energy cost to generate a pH gradient of two pH units is equal to $RT\ \ln([H^+_o]/[H^+_i])$, where the subscripts o and i stand for outside and inside the cell, respectively. At 25°C, the $\Delta G'$ for a 100-fold proton activity (pH 7.0 in versus pH 5.0 out) gradient is 2.7 kcal/mol.

Plasma membranes of all higher organisms contain enzymes that are embedded in the membrane that act as ion pumps. That is, they catalyze the transport of ions against their electrochemical potential. In physiology, transport that is thermodynamically uphill is termed active transport to distinguish it from the spontaneous flow of ions down their electrochemical potential. The energy needed

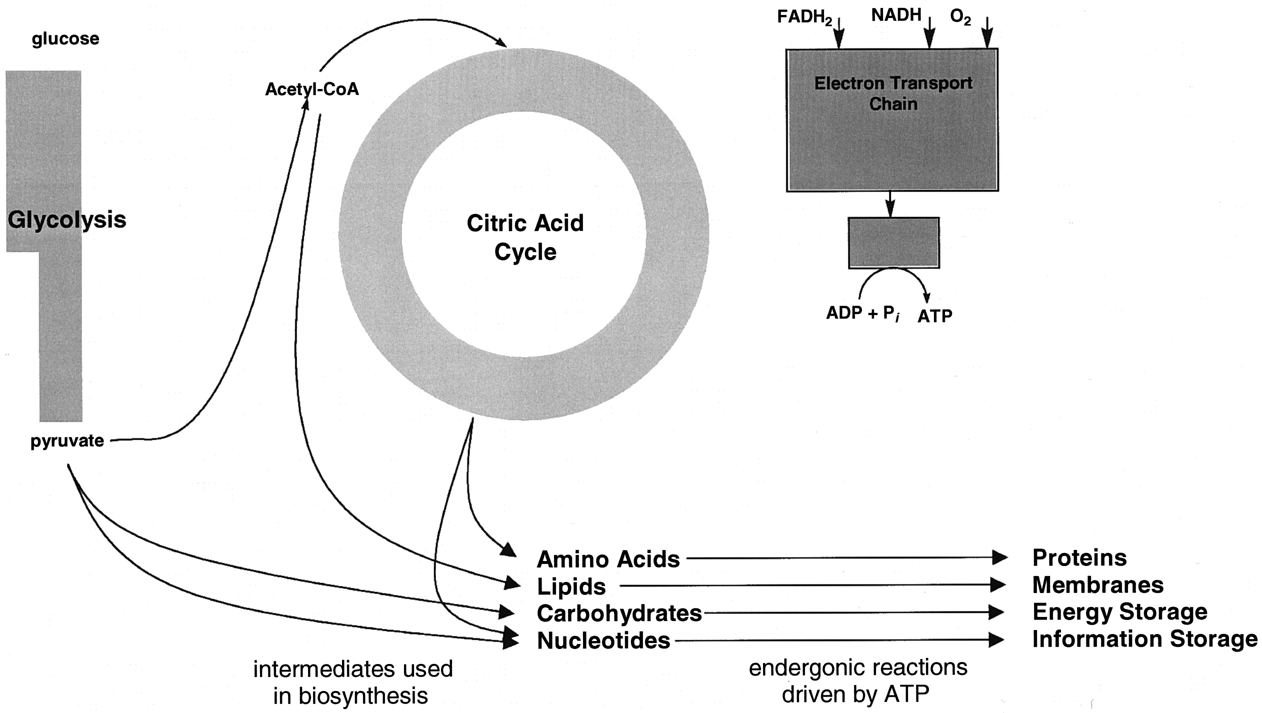


FIGURE 11 Uses of ATP. The diagram shows some of the major processes in cells that are powered by ATP hydrolysis.

for the active transport of ions across the plasma membrane is provided by the hydrolysis of ATP to ADP and P_i. As much as 75% of cellular ATP may be consumed simply to generate and maintain ion gradients.

The electrogenic ion pump in the plasma membrane of animal cells is the Na^+/K^+-ATPase. As shown in Fig. 12, three Na^+ ions are transported out of the cell and two K^+ ions are pumped in for each ATP that is hydrolyzed. Since three positively charged ions are exported, but only two imported, the Na^+/K^+-ATPase is electrogenic. The trans plasma membrane potential is on the order of −50 mV. In addition, the pump keeps the intracellular Na^+ concentration nearly 100-fold lower than that in the serum, and the intracellular concentration of K^+, about 30-fold higher than in serum.

Indirectly, the Na^+/K^+-ATPase provides the energy for the active transport of amino acids and some carbohydrates into cells. The plasma membrane contains specific proteins that mediate the transport of these molecules in a manner that is obligatorily linked to the cotransport of

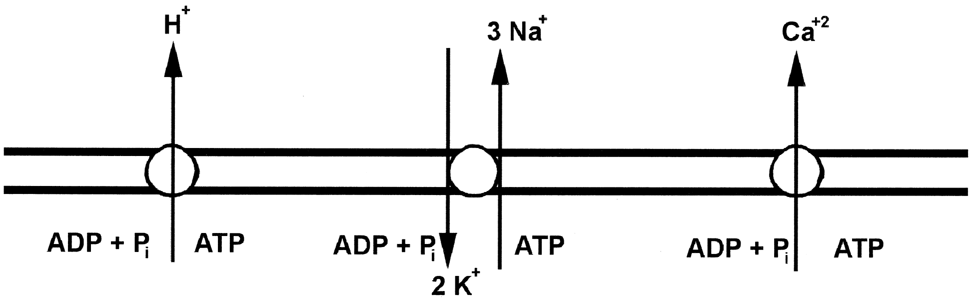


FIGURE 12 Some ion pumps in the plasma membrane. The Na^+/K^+-ATPase of animal cells uses the energy of ATP hydrolysis to move three Na^+ ions out of the cells and two K^+ ions in, which results in the generation of ion gradients and a membrane potential. Plant, yeast, and fungal cells do not have a Na^+/K^+-ATPase, but instead have a H^+-ATPase, as the electrogenic pump. The plasma membrane also contains a Ca^{2+}-ATPase that pumps Ca^{2+} out of cells to help keep the intracellular Ca^{2+} concentration low.

Na^+. Since the extracellular Na^+ concentration is higher than that in the cytoplasm and the membrane potential is negative, the Na^+ flows from outside to inside the cell. Assuming a membrane potential of −50 mV and a 100-fold Na^+ concentration gradient, the flow of Na^+ would liberate about 3.8 kcal/mol at 25°C. This exergonic flow of Na^+ provides the energy needed for the active transport of the amino acid or carbohydrate. Although Na^+ flux is the immediate source of energy for the active transport in Na^+-linked transporters, it is important to keep in mind that the ultimate energy source is ATP hydrolysis by the Na^+/K^+-ATPase.

Plants, yeasts, and fungi do not contain a Na^+/K^+-ATPase in their plasma membranes. Instead, they contain a H^+-ATPase that is the generator of the plasma membrane potential. The H^+-ATPase is structurally and mechanistically related to the Na^+/K^+-ATPase but translocates only H^+. The H^+-ATPase is capable of generating large electrochemical proton gradients. The imbalance in the Na^+ and K^+ concentrations between the inside and the outside of the plant cell is maintained by other mechanisms that include exchange transport of Na^+ for H^+.

The active transport of some organic molecules across the plasma membrane of plants, yeasts, and fungi is linked to the cotransport of H^+ down its eletrochemical gradient into the cell. An important example of proton-linked transport is that of sucrose loading into the vascular element, the phloem, that transports sucrose from the leaves to the remainder of a plant. The concentration of sucrose in phloem cells near leaves that are actively carrying out photosynthesis can be 0.5 M or higher, whereas that in the intracellular space, just 0.001 M. The energy cost of generating this gradient is 3.7 kcal/mol at 25°C. The immediate source of energy is proton flow, and the ultimate source, ATP hydrolysis by the H^+-ATPase.

The concentration of free Ca^{2+} (meaning that unbound to proteins and membrane lipids) in the cytoplasm of cells is normally maintained at a very low level. Under certain circumstances, however, transient increases in the cytoplasmic Ca^{2+} concentration are triggered. Ca^{2+} is a major player in the transmission of some hormonally induced signals in plants and animals. Muscle contraction is also induced by release of Ca^{2+} from internal membranes within muscle cells.

The plasma membrane contains an enzyme that catalyzes the export of Ca^{2+} from the cytoplasm at the expense of ATP hydrolysis. The Ca^{2+}-ATPase has features that place it in the category of plasma membrane enzymes that also includes the Na^+/K^+-ATPase and the H^+-ATPase. The Ca^{2+}-ATPase functions to keep the cytosolic Ca^{2+}concentration low ($<1\ \mu M$). It is not a major contributor to the generation of the membrane potential or to the energetics of the transport of bioorganic molecules.

Inhibitors of the enzyme responsible for the acidification of the stomach are well known and equally well-advertised alleviators of "heartburn." This enzyme is present in the parietal cells of the stomach and resembles the Na^+/K^+-ATPase. Instead of catalyzing the ATP-dependent exchange of Na^+ and K^+, the stomach acid pump excretes H^+ into the lumen of the stomach in exchange for K^+.

B. Biosynthetic Use of ATP

The input of energy in the form of the hydrolysis of ATP to either ADP and P_i or to adenosine monophosphate (AMP) and pyrophosphate powers the synthesis of biological molecules, including, as we have seen, carbohydrates in photosynthesis, proteins, DNA, RNA, and fatty acids. To delve into the role of ATP in biosynthesis in depth is not possible in this brief article. Aspects of fatty acid biosynthesis, however, reveal interesting principles of the energetics of biosynthetic pathways.

Fatty acids are oxidized completely to CO_2 and water by β-oxidation and the citric acid cycle. Acetyl CoA is the end product of β-oxidation of fatty acids and is the source of carbon for fatty acid biosynthesis. Yet, the pathways for fatty acid degradation and synthesis are so very different that they even occur within different compartments within cells. Fatty acid synthesis takes place in the cytoplasm of animal cells and in the plastids of plant cells, whereas β-oxidation is located in mitochondria in both animal and plant cells.

Often, the pathway for the synthesis of a compound differs significantly from that for its degradation. Among the reasons that the separation of synthetic and degradative pathways evolved are energetics and regulation. The oxidation of fatty acids to acetyl CoA is very exergonic. It is not feasible on energetic grounds to make fatty acids from acetyl CoA by reversing β-oxidation. Metabolism of carbohydrates and fats is regulated in mammals by a number of hormones, including insulin, glucagon, and epinephrine (adrenaline). Having separate pathways for the degradation and the biosynthesis makes it possible to turn off one pathway while up-regulating another. For example, glucagon and epinephrine selectively stimulate the breakdown of fats and fatty acids, whereas insulin has the opposite effect. The fine control of fatty acid metabolism that has evolved would clearly not be possible without the existence of separate pathways for biosynthesis and catabolism.

CO_2 is required for the synthesis of fatty acids. Yet, when fatty acid synthesis is carried out in the presence of radioactive CO_2, the fatty acid made is devoid of radioactivity. ATP is used to add CO_2 to a precursor, and in a subsequent step in the pathway of fatty acid biosynthesis,

this same CO_2 is released. This seemingly perplexing phenomenon may readily be explained on an energetic basis.

Acetyl CoA is carboxylated by using bicarbonate as the source of CO_2 and ATP hydrolysis as the source of energy:

$$\text{acetyl CoA} + \text{ATP} + HCO_3^- \rightarrow \text{malonyl CoA} + \text{ADP} + P_i. \quad (12)$$

The enzyme that catalyzes this reaction, acetyl CoA carboxylase, contains biotin, one of the B vitamins. Several other vitamins, including niacin (part of NAD^+ and $NADP^+$) and riboflavin (part of FAD), are essential players in metabolism.

The carboxylation of acetyl CoA without the hydrolysis of ATP is energetically unfavorable. The exergonic hydrolysis of ATP pulls the reaction toward malonyl CoA synthesis. But why bother to carboxylate acetyl CoA?

All the carbon atoms in synthesized fatty acids are derived from the acetyl group of acetyl CoA. In principle, fatty acids could be made by condensation of acetyl units and subsequent reduction. However, the condensation of two acetyl CoA molecules is energetically unfavorable. The release of CO_2 as part of a reaction helps to drive a reaction to completion. The oxidative decarboxylation reactions of the citric acid cycle illustrate this fact. The loss of CO_2 from the malonyl group as it condenses with the acetyl group bound to the fatty acid synthetase drives the condensation reaction. The resulting β-keto compound is reduced to the level of a hydrocarbon by NADPH.

ATP hydrolysis provided the energy for the carboxylation of acetyl CoA. The immediate energy source for the condensation reaction was the loss of the same CO_2 molecule added to the acetyl CoA. It is clear that CO_2 plays a catalytic but essential role in fatty acid biosynthesis.

C. ATP and Motility

At macroscopic and microscopic levels, ATP hydrolysis results in movements. The most familiar of these movements are those caused by muscle contraction. Muscle contraction is an example of the conversion of the phosphate bond (chemical) energy of ATP to mechanical energy. Vertebrate muscle is composed of two types of filaments, thick and thin. The protein myosin is the major component of the thick filaments, whereas actin and other proteins make up the thin filaments. The thick and the thin filaments are interdigitated. Muscle contraction is thought to take place by a sliding of the thin filaments relative to the thick filaments. Myosin has ATPase activity. The catalytic site in myosin is located on a part of the molecule (the head) that interacts with the actin filaments. ATP hydrolysis is thought to cause changes in the interactions of the myosin head with the actin filaments such that the head moves along the actin filament in one direction.

Muscle contraction is regulated by a Ca^{2+}-binding protein in the thin filaments. Ca^{2+} is required for muscle contraction. During rest, the concentration of Ca^{2+} in muscle cells is kept low by the operation of two Ca^{2+}-ATPases, one in the plasma membrane and the other in internal membranes called the sarcoplasmic reticulum. The release of Ca^{2+} triggers muscle contraction, and its uptake into the lumen of the sarcoplasmic reticulum causes relaxation.

V. CONCLUDING STATEMENTS

There are two aspects of bioenergetics that we want to emphasize at the end of this article. These are the dependence of life on photosynthesis and the diversity of energy interconversions in living systems.

Photosynthesis is the only major biological process that uses a source of energy, sunlight, from outside the earth's environment to convert inorganic molecules to organic molecules, including carbohydrates, proteins, nucleic acids, lipids, and pigments. Green plants and algae are autotrophs; they make their own food. Actually, plants synthesize all the thousands of compounds that they contain from CO_2, H_2O, and inorganic nitrogen and sulfur compounds absorbed through the roots. The only source of carbon is CO_2, which is assimilated through photosynthesis. Most other organisms are heterotrophs; they must take up and catabolize carbohydrates and fats to provide the energy to sustain life. The ultimate source of these compounds is photosynthesis, and the source of energy for their synthesis, sunlight. All heterotrophic organisms are dependent upon photosynthesis for their existence.

Animals also depend on plants for essential organic molecules that they are unable to make. We call some of these molecules vitamins. Several vitamins, including niacin, riboflavin, pyridoxine, and biotin, are key players in catabolic and anabolic metabolism, and deficiencies in these vitamins have severe effects. Also, animals are incapable of synthesizing polyunsaturated fatty acids (fatty acids with more than one double bond). Polyunsaturated fatty acids are essential components of membrane lipids and must be obtained in the diet. So, the next time you have a salad, pay a tribute to photosynthesis.

In photosynthesis, the electromagnetic energy of light is converted to chemical energy in the form of organic molecules. The primary photochemical reactions are electron transfer reactions that create oxidized chlorophylls and reduced acceptors. The reaction center chlorophylls and the acceptors are arranged within the photosynthetic membrane so that the electrons are transferred at least partway across the membrane. Thus, the membrane is charged by the primary electron transport, and electrical work has been done. The electron transport that follows the primary

reactions is directly linked to the transmembrane flow of protons into the lumen of the membrane. This proton flow results in the generation of an electrochemical proton gradient. Essentially, part of the light energy is conserved by formation of this gradient as well as by formation of the strong reducing agent NADPH. The flow of protons provides the energy needed for the synthesis of the terminal phosphate anyhydride bond of ATP, an example of the conversion of the osmotic and electrical energy of the proton gradient to chemical bond energy. The syntheses of ATP and NADPH capture some of the light energy. In turn, ATP and NADPH drive the unfavorable reduction of CO_2 by H_2O to form carbohydrates and O_2.

Organisms, especially bacteria, have evolved novel bioenergetic mechanisms that are well suited to their environments. For example, the bacterium *Halobacter halobium* lives in salt marshes and requires NaCl at concentrations that kill other organisms. These halophilic bacteria contain patches of a purple protein, halorhodopsin, on its plasma membrane. Halorhodopsin is a light-driven proton pump and its operation causes protons to be ejected from the cells. The resulting electrochemical proton gradient may be used to drive ATP synthesis or the transport of biochemicals. Given the diversity of the environments in which organisms grow, it is possible that biochemists will uncover new ways in which organisms meet their energetic needs. Perhaps future bioenergeticists will have the opportunity to unravel the mysteries of organisms from planets other than Earth.

SEE ALSO THE FOLLOWING ARTICLES

Carbohydrates • Carbon Cycle • Chromatin Structure and Modification • Electron Transfer Reactions • Energy Flows in Ecology and in the Economy • Energy Transfer, Intramolecular • Ion Transport Across Biological Membranes • Lipoprotein/Cholesterol Metabolism • Protein Synthesis • Thermodynamics

BIBLIOGRAPHY

Cramer, W. A., and Knaff, D. B. (1990). "Energy Transduction in Biological Membranes: A Text of Bioenergetics," Springer-Verlag, New York.

Garrett, R. H., and Grisham, C. M. (1999). "Biochemistry," 2nd ed., Saunders College Publishing, Fort Worth.

McCarty, R. E. (1999). "Chemiosmotic Coupling," *In* "Encyclopedia of Molecular Biology" (T. Creighton, ed.), pp. 402–408, John Wiley and Sons, Inc., New York.

Nichols, D. G., and Ferguson, S. J. (1992). "Bioenergetics 2," Academic Press, London.

Ort, D. R., and Yocum, C. F., eds. (1996). "Oxgenic Photosynthesis: The Light Reactions," Kluwer Academic, Dordrechtshill, Norwell, MA.

Enzyme Mechanisms

Stephen J. Benkovic
Ann M. Valentine
Pennsylvania State University

I. Introduction to Enzymes as Catalysts
II. Enzyme Kinetics
III. Illustrative Examples
IV. Origins of the Catalytic Efficiency of Enzymes

GLOSSARY

Inhibitor A molecule that by binding to the enzyme lowers its activity (i.e., its ability to process the substrate).

Intermediate A molecular species usually bound to the enzyme that exists transiently in the course of converting the substrate of the enzyme to its product.

Product A molecule that results from a chemical transformation of its precursor substrate at an enzyme's active site.

Substrate A molecule that binds to an enzyme's active site and is chemically transformed.

THE IMPETUS for understanding how enzymes function is inspired by their enormous catalytic efficiency and their exquisite substrate stereospecificity. With the advent of the determination of enzyme structure and the application of physical organic tools to examine the reaction coordinate for the enzymatic transformation of the substrate, penetrating insights have been gained as to the number and magnitude of the kinetic steps in the catalytic cycle, the chemical nature of intermediates, and the function of the active site residues contributed by the enzyme. This noninclusive article describes a small number of enzyme-catalyzed reactions from the viewpoint of protein structure, reaction kinetics, and probable chemical identity of intermediates along the reaction pathway. It concludes with our musings as to the chemical origins of the unusual catalytic properties of enzymes.

I. INTRODUCTION TO ENZYMES AS CATALYSTS

Enzymes are biological molecules which accelerate the rate, and often direct the specificity, of a chemical reaction. Like all catalysts, they are not themselves consumed in the reactions in which they participate but are regenerated to take part in multiple cycles. Transformations which are very slow, such as the breakdown of DNA, can be accelerated by many orders of magnitude by an appropriate enzyme. Enzymes cannot catalyze reactions that are not thermodynamically favorable, but they can facilitate and accelerate those that are favorable but slow and can couple unfavorable reactions to even more favorable ones. Most enzymes are proteins and thus are made up of amino acids. Recently it was discovered that RNA molecules can

also be enzymes; this type of enzyme activity will not be discussed in this article. For the purposes of this discussion, we will explore how protein molecules, sometimes in conjunction with cofactors, use chemistry to convert substrates to products. The availability within a protein, or through cofactors, of nucleophiles or electrophiles, acid or base residues, redox centers, or other features associated with chemical catalysts, when coupled to the selective pressure of evolution, has afforded selective and efficient catalysts. The study of enzyme mechanisms aims to define as precisely as possible the nature of the chemical steps that effect these conversions.

A logical starting point is to consider the structures of four of the representative enzymes depicted in this article: chymotrypsin, dihydrofolate reductase, aspartate aminotransferase, and cytochrome P450. In general, one observes a well-defined binding site for capturing the substrate and executing the chemical transformation through various polar and nonpolar interactions between the substrate and the amino acids that line the active site. Much of the rate acceleration (up to 10^8-fold) for enzymatic catalysis can often be attributed to the juxtaposition of substrates and catalytic residues within the active site cavity. There are currently available X-ray crystallographic structures of enzymes, many with active sites occupied with inhibitors and determined to a resolution of less than 2.5 Å, which permit inferences as to the mechanism of the chemical transformation. Nuclear magnetic resonance (NMR) and optical spectroscopic methods provide important, complementary data on solution structure. Despite considerable differences in the primary amino acid sequence, the overall protein fold with its α-helical and β-sheet secondary structural elements is often retained for classes of transformations that are related through a common mechanistic species and thus constitute members of a protein superfamily. One implication is that the entire tertiary structure, not merely the active site, is important in the efficiency and selectivity of the chemical transformation. The structures we have chosen will serve to illustrate how the convergence of the knowledge of structure with the output from other experimental tools provides arguments for probable mechanisms of catalysis.

II. ENZYME KINETICS

The study of the rates of enzyme-catalyzed transformations provides invaluable information as to the number of steps and their magnitude in the catalytic process. The most common method is to use steady-state conditions in which the enzyme is at $<10^{-8} M$ concentration and the substrate(s) μM or higher. In the simplest case of the conversion of a single substrate to product, the kinetic scheme is generally represented by

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P,$$

where S and P are the substrate and the product, E is the free enzyme, and ES is the associated enzyme · substrate complex, also called a Michaelis complex. The rate constants k_1, k_{-1}, k_2, and k_{-2} describe the rates of each step in the reaction. Because the concentration of ES is not changing, and so is at the steady state, the kinetic scheme can be solved by relating the initial velocity at a given substrate concentration to both the maximum velocity, $V_{\max}$, and the substrate concentration at which the initial velocity reaches one-half the maximum velocity, K_M, through the equation

$$V = V_{\max}/(1 + K_M/[S]).$$

The term K_M is the ratio $(k_{-1} + k_2)/k_1$ and only approximates the binding of S to E. The turnover number, or k_{cat}, is simply $V_{\max}/[E_o]$ where E_o is the total enzyme concentration. A description of the transformation of substrate to product generally shows V as a hyperbolic function of S concentration with V increasing asymptotically toward $V_{\max}$ as the active site becomes saturated with S.

Even in this simple case, the extraction of the magnitude of the four specific rate constants requires numerical analysis, with additional complexity being introduced by the appearance of intermediates or the requirement for a second or third substrate. These complications lead to equations in which the additional rate constants cannot be calculated from steady-state data. However, the analogous terms for K_M and k_{cat} can be calculated and hold similar meanings. Perhaps the most useful application of steady-state kinetics at this level is the recognition of diagnostic patterns in the reciprocal replots of the initial velocity data as a function of substrate concentration. Two-substrate reactions fall into two general classes represented by

$$E + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EA \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EX + C$$

$$EX + B \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} EXB \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} E + D$$

and

$$E + A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EA + B \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EAB \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} EDC \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} ED$$

$$+ C \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} E + D.$$

The difference is that in the former process a fragment X of substrate A is transferred covalently to the enzyme and

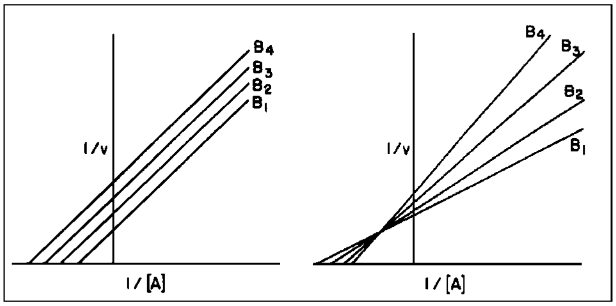
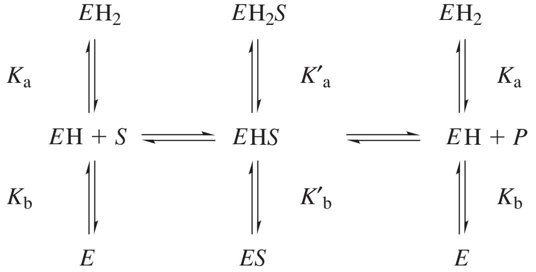


FIGURE 1 Plots of the reciprocal initial velocity against the reciprocal concentration of substrate *A* for a two-substrate reaction at several different concentrations of substrate *B*. The plot on the left reflects a mechanism in which a free enzyme bearing a covalently linked group is generated, while that on the right shows a sequential one in which two substrates bind and the reaction occurs. [From Hammes, G. G. (1982). Enzyme Catalysis and Regulation. Academic Press, New York. Used with permission.]

then to the second substrate, whereas in the latter no free enzyme bearing a covalently linked fragment X is formed. Within the second pathway, the addition of A and B, and similarly the release of products C and D, can be ordered (as written) or random. The two pathways give rise to the representative graphs shown in Fig. 1.

As one might imagine, the kinetic rate laws associated with these mechanisms are generally too complex for dissection of single steps or their evaluation. Moreover, the method provides evidence for only the minimal number of intermediates in a pathway since the form of the equations is unchanged by including multiple species.

Steady-state kinetic parameters such as k_{cat} and K_M can vary when they are studied as a function of pH. After one corrects for ionizations of the substrate and controls for possible effects on the native structure of the enzyme, variations in k_{cat} and K_M can often be assigned to ionizations of acid/base groups at the active site of the enzyme. The term k_{cat}/K_M reflects the proton dissociation constants of the free enzyme, provided that the proton transfers remain fast relative to other steps in the pathway. In the simple one-intermediate kinetic sequence expanded to implicate two ionizations, the term k_{cat}/K_M would display pK_a and pK_b; the term k_{cat} would reflect pK'_a and pK'_b. The pH dependence of the k_{cat} parameter affords information about the substrate-bound state.



Materials that bind to the enzyme either at the active site or at a distal site and slow the turnover of the enzyme but are not themselves transformed act as inhibitors. These compounds may or may not be structurally similar to the substrate; nevertheless, their binding, particularly at the active site, often provides important complexes for structure determination. The most commonly studied type of inhibition is termed competitive, which means that the substrate and the inhibitor compete directly for the active site of the enzyme. The effect of this type of inhibitor on the steady-state kinetic parameters is to alter the graphical evaluation of the Michaelis constant but not the value of V_{max}, which can still be attained in the presence of the inhibitor provided that the substrate concentration is high enough. Binding of the inhibitor to regions divorced from that binding the substrate always affects the evaluation of V_{max} because no concentration of substrate is sufficient to displace the inhibitor.

The most useful approaches for obtaining information regarding the existence of intermediates and their lifetimes are fast reaction methods that mix enzyme and substrate within milliseconds, which permits the observation of single turnover events by various spectroscopic methods. Alternatively the reaction is rapidly quenched at known time intervals and its progress is analyzed chromatographically. In many cases in which an intermediate accumulates to the level of the enzyme concentration, such methods reveal the presence of "burst kinetic" that feature the rapid buildup of the intermediate in the transient phase followed by its slower rate of formation/decay in the steady state. The simplest kinetic scheme consistent with this phenomenon is given by

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P,$$

where the rate constants are in the order $k_1[S] > k_2 > k_3$. The amplitude of the burst can provide the concentration of active sites in an enzyme preparation. By varying the concentration of S, one can find values for k_{-1}/k_1, k_2, and k_3. There are many variations on transient kinetics, as will be illustrated in our case studies of individual enzymes.

III. ILLUSTRATIVE EXAMPLES

A. α-Chymotrypsin

Alpha-chymotrypsin (Fig. 2) catalyzes the facile hydrolysis of peptide bonds, in particular those adjacent to the carboxyl group of aromatic amino acids (tryptophan, tyrosine, phenylalanine) as well as a variety of esters derived from similar N-acylated amino acids. The enzyme

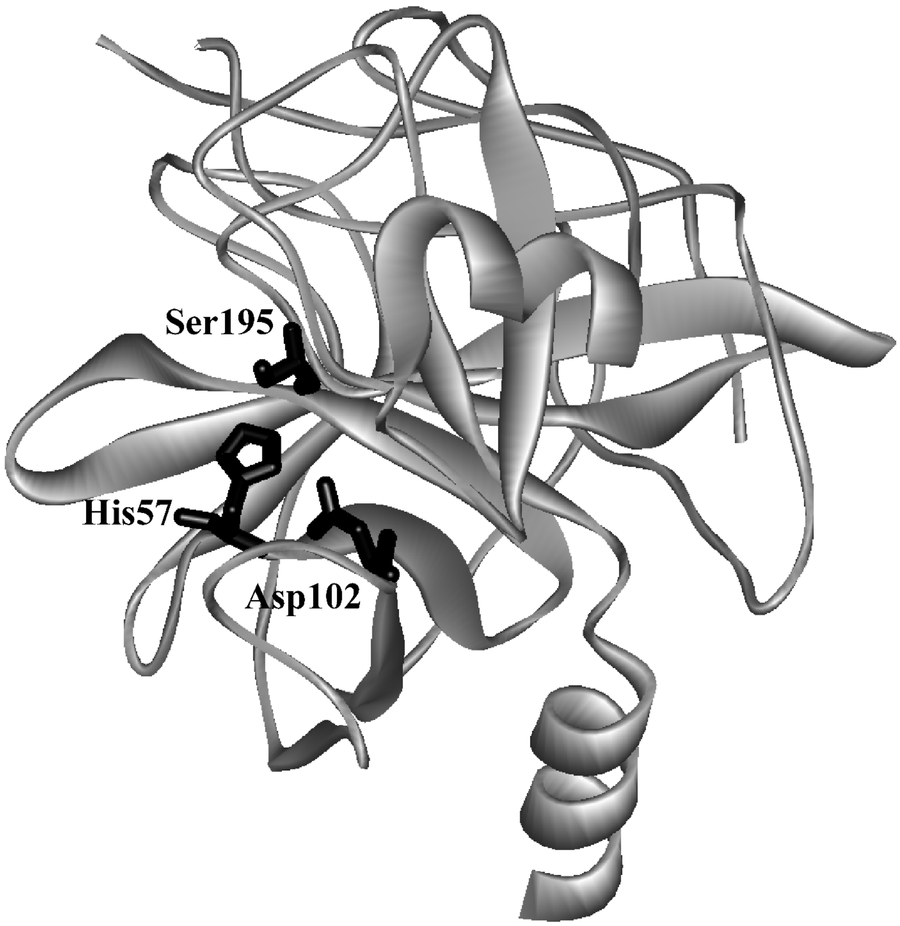


FIGURE 2 The crystal structure of α-chymotrypsin showing the catalytic triad of amino acid side chains. [Adapted from Blevins, R. A., and Tulinsky, A. (1985). "The refinement and crystal structure of the dimer of α-chymotrypsin at 1.67 Å resolution," *J. Biol. Chem.* **260,** 4264–4275.]

has been the subject of intensive mechanistic study, most of which occurred well before a crystal structure was available.

A key insight was provided by studying the enzyme-catalyzed hydrolysis of *p*-nitrophenyl acetate. Transient kinetic studies revealed burst kinetics (Fig. 3) with an initial rapid liberation of *p*-nitrophenolate followed by a slower steady-state rate. The biphasic time course is consistent with the existence of two intermediates (*ES* and acyl-*E*), with the second accumulating owing to its slower breakdown to product. The intermediate is a covalent enzyme species acylated at serine-195 (see Fig. 2), a fact initially revealed by chemically esterifying this enzyme residue specifically and irreversibly with diisopropylphosphorofluoridate. No burst kinetics is seen with amide substrates because the acylation step limits turnover. The same intermediate, however, is formed as shown by partitioning experiments in which an exogenous nucleophile such as hydroxylamine is added to compete with water in the deacylation step. The result revealed equivalent levels of hydroxamate and acid products formed from either amide or ester substrates derived from a common amino acid, which implicated the presence of the intermediate in both enzyme-catalyzed processes.

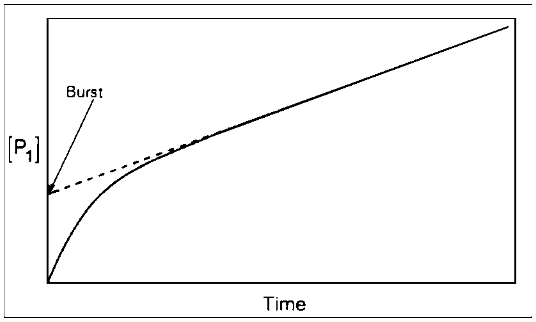


FIGURE 3 Plot of the burst in hydrolysis of *p*-nitrophenyl acetate. The concentration of product is observed as a function of time. [From Fersht, A. (1999). Structure and Mechanism in Protein Science. W. H. Freeman and Company, New York. Used with permission.]

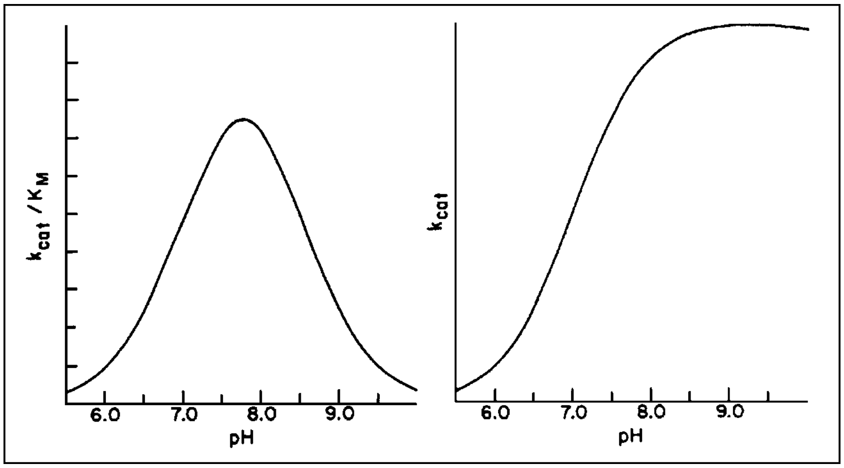


FIGURE 4 The pH dependence of k_{cat}/K_M and k_{cat} for the α-chymotrypsin-catalyzed hydrolysis of esters and amides. [From Hammes, G. G. (1982). Enzyme Catalysis and Regulation. Academic Press, New York. Used with permission.]

The pH dependence of the steady-state kinetic parameters is shown in Fig. 4 and implicates the ionization of two groups in the free enzyme and one in the *ES* complex. These data combined again with chemical modification studies (now superseded by site-specific mutagenesis) implicated histidine-57 (p$K_a \sim 7$) and the *N*-terminal amino acid isoleucine (p$K_a \sim 8.5$). The latter forms a salt bridge with aspartate-194 that helps maintain the active structure of the enzyme; the former is involved in general acid–base chemistry at the active site.

These data, along with further information derived from the reaction of specific substrates with the enzyme by using stopped-flow methods, led to the elucidation of a kinetic sequence that consistently implicated the acylation and deacylation of Ser195 assisted by His57 and Asp102. The crystal structure of chymotrypsin (Fig. 2) reveals that these three residues form a catalytic triad, a feature repeated for many hydrolytic enzymes. This triad operates within a well-defined binding site that is lined with nonpolar amino acids capable of van der Waals interactions with polypeptide substrates containing aromatic side chains. A plausible mechanism is outlined in Fig. 5 in terms of the chemistry occurring during the individual kinetic steps.

The key features of this mechanism require the participation of the serine hydroxyl as a nucleophile whose attack on the carbonyl of the substrate is facilitated through proton abstraction by the imidazole nitrogen of His57 and its redonation to the amine-leaving group. Deacylation of the enzyme follows general base catalysis of water attack again by His57 and the return of the enzyme to its resting state. Catalysis of the chemical process through the participation of the side chains of an enzyme in proton, hydride, and electron transfer is a hallmark of enzyme catalysis and can occur efficiently in the confines of the active site owing to the optimal alignment and juxtapositioning of the substrate for chemical reaction.

B. Dihydrofolate Reductase

Dihydrofolate reductase (DHFR) catalyzes the reduction of 7,8-dihydrofolate (H_2F) by nicotinamide adenine dinucleotide phosphate (reduced form) (NADPH) to form 5,6,7,8-tetrahydrofolate (H_4F), a key step in furnishing the parental cofactor needed for *de novo* pyrimidine and purine biosynthesis. The enzyme has been the target of antitumor and antimicrobial drugs. A complete kinetic scheme (Fig. 6) obtained primarily through transient kinetics has been described for the enzyme from *Escherichia coli* as well as other sources and provides a second case study as to how to define the catalytic process.

Measurement of the rates of binding and dissociation of substrate and cofactors provided valuable insights into the identity of rate-limiting kinetic steps in the scheme shown in Fig. 6. Two procedures were used. In the first, direct observation of changes in the intrinsic enzyme or NADPH fluorescence upon ligand binding showed that the addition of ligand was biphasic in accord with the existence of two conformers, of which only one bound the ligand:

$$\text{DHFR}_1 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \text{DHFR}_2 + L \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{DHFR}_2 \cdot L.$$

The rate of the initial fast phase and its amplitude are associated with the binding of L to DHFR_2 (k_1, k_{-1}) and the level of DHFR_2; the rate of the second phase is the conversion of DHFR_1 to DHFR_2 (k_2). The method was extended to the binding of a second ligand to binary

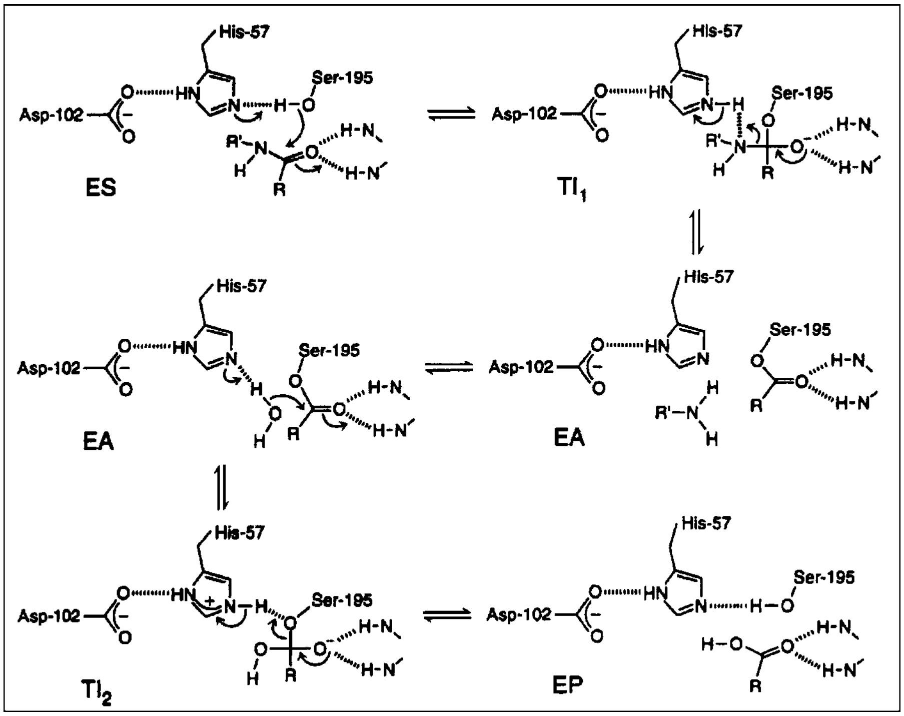


FIGURE 5 The mechanism of amide hydrolysis by α-chymotrypsin. [From Fersht, A. (1999). Structure and Mechanism in Protein Science. W. H. Freeman and Company, New York. Used with permission.]

$DHFR_2 \cdot L$ complexes and revealed that the binding of various ligands was near the diffusion-controlled limit.

In the second procedure a competitive trapping technique was employed in which the enzyme–ligand complex is mixed with an excess of a second ligand that competes for the binding site. With this method, k_{-1} is measured accurately when $k_T[T] \gg k_1[L_1]$, k_{-1}, and k_{-T}. This

$$\mathrm{DHFR} \cdot L_1 \underset{k_1}{\overset{k_{-1}}{\rightleftharpoons}} \mathrm{DHFR} + L_1 \underset{k_{-T}}{\overset{k_T[T]}{\rightleftharpoons}} \mathrm{DHFR} \cdot T.$$

procedure identified a preferred pathway for dissociation of the product H_4F as the rate-limiting step in the steady-state cycle. The assistance of the cofactor NADPH in promoting product dissociation is an unusual feature, though not limited to DHFR, and follows the rapid loss of $NADP^+$.

Events around the chemical step of reduction/oxidation were monitored by directly observing the conversion of NADPH to $NADP^+$. The kinetics are again biphasic owing to the rapidity of the hydride transfer process; that the rapid phase is associated with the chemical step is verified by the observation of a kinetic deuterium isotope effect of 3 when the transferring hydrogen of the NADPH is replaced with deuterium. This step shows a pH dependence with a pK_a of 6.5 that implicates the Asp125 (27 in *E. coli*) in the proton transfer events required to complete the reduction.

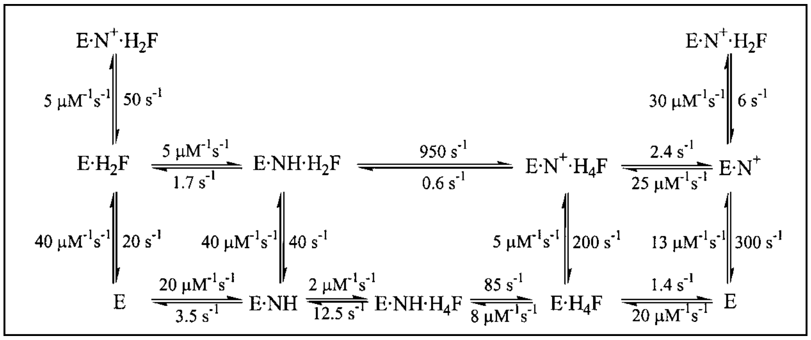


FIGURE 6 The kinetic scheme for conversion of H_2F to H_4F by DHFR, including the rate constants for each step at 25°C. In this scheme, NH represents NADPH and N^+ represents $NADP^+$.

Measurement of this step in the reverse direction (i.e., for DHFR · $NADP^+$ · H_4F) coupled with determination of the overall equilibrium constant permitted construction of Fig. 6.

The kinetic scheme served as the basis for the explanation of the contribution of various elements of the protein to its function. Site-specific mutagenesis is a technique in which one or more amino acids are replaced by other amino acids through alteration of the gene encoding the enzyme. For the mutant proteins, the same kinetic scheme was reconstructed to calculate the free energy differences arising from changes in the kinetic steps caused by the mutations. Replacing the hydrophobic residues such as Phe30 and Leu54 (Fig. 7) singly or pairwise with other amino acids revealed that the cumulative effect of two mutations was generally nonadditive in terms of the free energy associated with individual steps in Fig. 6, consistent with long-range interactions across the enzyme active site mediated by bound substrate and cofactor. The nonadditivity differed for each step in Fig. 6, which implicated differing conformations of the protein as arising throughout the catalytic cycle.

Of particular interest was the discovery that changes in the amino acid sequence at loci outside the active site also strongly influence (by a factor of $>10^2$) the rate of the chemical step. In combination with dynamic NMR measurements and molecular mechanics calculations, this observation has been attributed to the importance for catalysis of long-range motions that occur across the entire DHFR protein molecule. The protein fold through its complex vibrational modes apparently may couple some set of motions to a promotional vibration that fosters passage of the reactive ternary complex over the activation barrier.

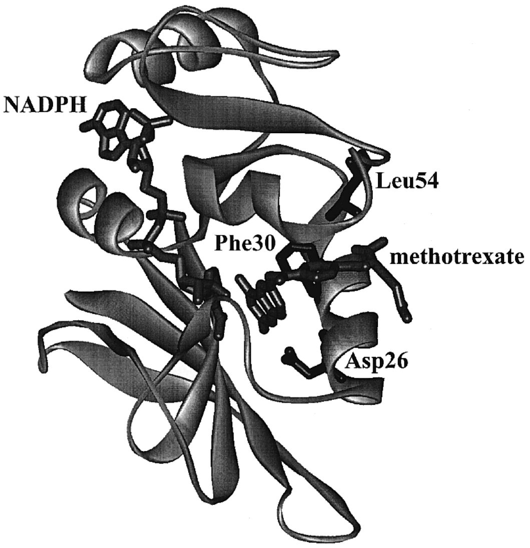


FIGURE 7 Crystal structure of DHFR from *Lactobacillus casei* with methotrexate (a strong inhibitor) and NADPH bound. Amino acid residues discussed in the text are labeled. [Adapted from Bolin, J. T. *et al.* (1982). "Crystal structures of *Escherichia coli and Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution," *J. Biol. Chem.* **257,** 13650–13662.]

C. Phosphate Transfer

Enzymes that catalyze the transfer of a phosphoryl moiety between two substrates have provided excellent examples of the use of isotopes in kinetic and stereochemical studies. The enzyme hexokinase, which promotes the conversion of glucose plus ATP to glucose-6-phosphate and ADP has been the subject of kinetic studies that suggested an ordered kinetic sequence with glucose being the first substrate to add and glucose-6-P the last product to be released. Specific information on the identity of rate-limiting steps and the steady-state levels of reaction intermediates was obtained by isotope trapping studies. In its simplest form, enzyme and isotopically labeled substrate (S^*) are incubated (the pulse) and rapidly diluted into excess unlabeled substrate (the chase), and allowed to react for a chosen time. Then the reaction is stopped by a quenching reagent that jumps the pH or denatures the enzyme. From the amount of $E \cdot S^*$ converted to product versus that lost to dissociation (replacement by S gives nonlabeled product) the dissociation rate of S^* from E and other ES complexes can be calculated.

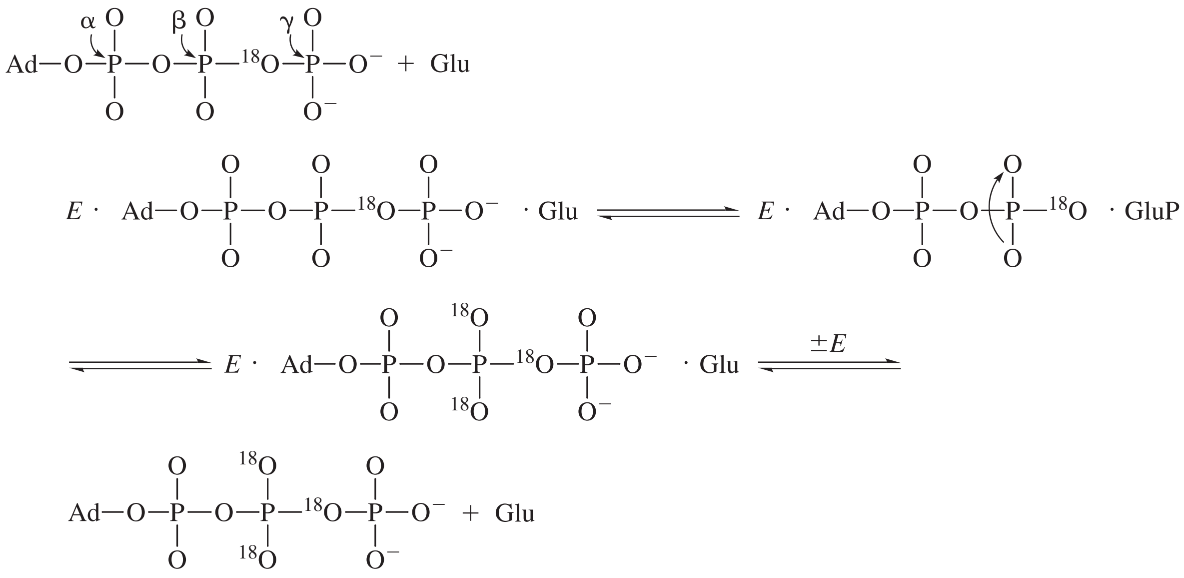
This method has been used in the study of the partitioning of ES complexes in the steady state. In the case of hexokinase, the question was the partitioning of the functional $E \cdot$ glucose · ATP complex between product formation and substrate release. For glucose the relevant scheme is

$$E + \text{Glc}^* \xleftarrow{k_{\text{off}}^{\text{Glc}^*}} E \cdot \text{Glc}^* \xleftarrow{k_{\text{off}}^{\text{ATP}}} E \cdot \text{Glc}^* \cdot \text{ATP} \underset{k_{-\text{c}}}{\overset{k_{\text{c}}}{\rightleftharpoons}}$$

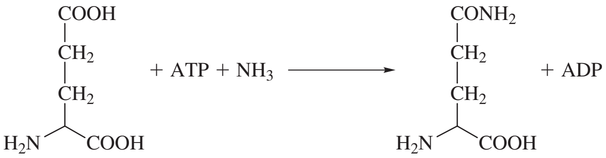
$$E \cdot \text{Glc}^*\text{-6-P} \cdot \text{ADP} \xrightarrow{k_{\text{off}}^{\text{ADP}}} E \cdot \text{Glc}^*\text{-6-P} + \text{ADP} \longrightarrow E \cdot$$

In this case the reaction is allowed to reach steady-state turnover, and the solution is either stopped by quench or chased by addition of excess unlabeled substrate followed by a delay sufficient for several turnovers then addition of quench. The presence of a difference in the level of the labeled product obtained by the two procedures represents the concentration of $E \cdot \text{Glc} \cdot \text{ATP}^*$ complex in the steady state, which is approximately 50% of E_{T}, the total enzyme concentration. The observed steady-state and pretransient rates are consistent with steps k_{c} and $k_{-\text{c}}$ being at equilibrium relative to $k_{\text{off}}^{\text{ADP}}$, which is typical for many phosphotransfer enzymes in which the chemical steps are generally not rate limiting. Additional information can be obtained by using the label in the second substrate (i.e., [γ-^{32}P]ATP) and following a similar protocol, which thereby allows calculation of the dissociation rate of ATP

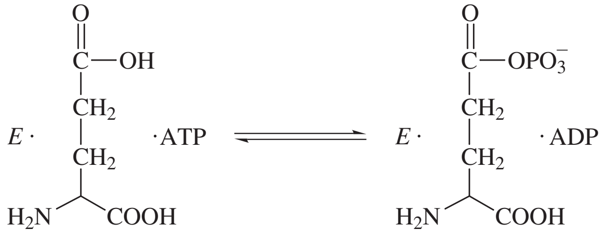
from $E \cdot \text{Glc} \cdot \text{ATP}$. In this case $E \cdot \text{Glc} \cdot \text{ATP}^*$ is approximately 25% of E_T, which requires that k_{off}^{ATP} compete with the dissociation of ADP (k_{off}^{ADP}) from $E \cdot \text{Glc-6-P} \cdot \text{ADP}$. In this manner the individual rate constants for hexokinase were largely determined and the order of substrate association was verified.



Isotopic labeling has also been cleverly used to demonstrate the existence of enzyme-bound intermediates that do not readily dissociate into solution. The enzyme glutamine synthetase catalyzes the formation of glutamine from ATP and ammonia possibly through a tightly bound glutamyl phosphate intermediate.



If the formation of glutamyl phosphate were reversible and occurred in the absence of ammonia, then the presence of a symmetric torsion motion at the cleavage site might be used to



detect an isotopic exchange brought about by glutamyl phosphate formation. The synthesis of γ-$^{18}O_4$P-ATP containing an ^{18}O label in the $\beta\gamma$ bridging oxygen provides the necessary probe for finding this intermediate by means of the process below. In these experiments, isotopes are used as labels so that the fate of a particular atom may be followed throughout the course of the reaction.

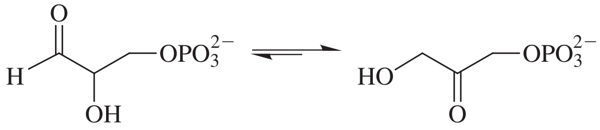
The appearance of ^{18}O in the nonbridging oxygens of the β-phosphate can be measured by mass spectrometric and NMR methods. The extent of equilibration is partially inhibited by the presence of ammonia as required if glutamyl phosphate is a reaction intermediate.

Isotopic labeling studies of phosphotranferase reactions culminated in the synthesis of ATP chiral at the γ-phosphorus. Chirality was achieved by the synthesis of [γ-^{16}O,^{17}O,^{18}O]ATP of one configuration, and the analysis of its chirality was achieved by stereochemically controlled transfer of the γ-phosphoryl moiety to (S)-propane-1,2-diol where the absolute configuration was determined by a chemical/mass spectrometric sequence. The observation of inversion of configuration has been accepted as evidence of an "in-line" displacement mechanism at phosphorus by the two bound substrates; the observation of retention of configuration was used to implicate the existence of a phosphoryl enzyme intermediate in the phosphoryl transfer process. For hexokinase, our case study, the finding is one of inversion, consistent with a direct transfer mechanism.

D. Triosephosphate Isomerase

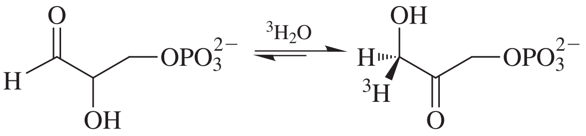
Triosephosphate isomerase (TIM) catalyzes the interconversion of D-glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP). The equilibrium

lies far to the side of DHAP, hence the longer arrow pointing to that compound. The enzyme operates with a turnover number of $\sim 10^7\ s^{-1}$, which is nearly as fast as the diffusion-controlled limit. TIM is therefore called an almost perfectly evolved enzyme because no catalytic refinement could make the rate faster than it already is.

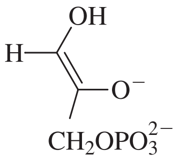


G3P DHAP

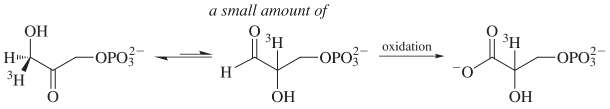
Many tools have been used to study the TIM mechanism, including X-ray crystallography and NMR, site-directed mutagenesis, and affinity labeling. Strong evidence for the mechanism, however, was supplied by studies using isotopic labeling of substrates. It was found that if the above reaction was carried out in tritiated water, one atom of tritium was stereospecifically incorporated into DHAP.



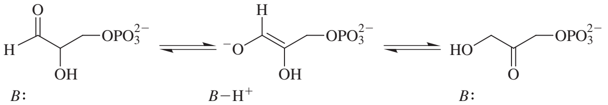
This result suggested that a base abstracts a proton from the substrate, and the proton then undergoes exchange with labeled protons from the solvent before being added back to form the product stereospecifically. The existence of a *cis*-enediol intermediate (shown below) would account for these observations, if the enzyme added the proton back to the same face of the enediol that it was abstracted from.



One unresolved question was whether only a single protein base was involved (so that the transfer was from substrate to base and directly back to form product) or whether a different base was responsible for protonation as part of a more extensive proton relay. The nature of the protein base was explored by doing a similar experiment to the one described above but in the other direction; that is, by labeling the DHAP and observing its conversion to G3P. Although the equilibrium lies far to the side of the DHAP, trapping by irreversible oxidation by G3P dehydrogenase of any G3P formed was used to convert significant quantities of DHAP. If the DHAP was labeled at C1, a small but measurable amount of the label was transferred to C2.



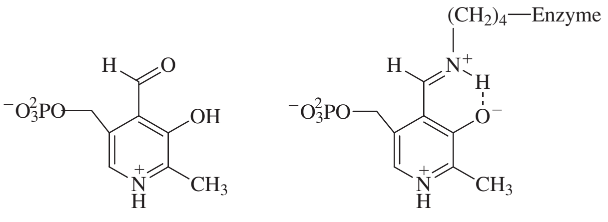
This result suggested that a single base was involved in the proton abstraction/proton addition step. If more than one base were involved, the chance that any label would not be washed out by the solvent and would be added to the deprotonated intermediate would be vanishingly small. In combination with other kinds of experiments, isotopic labeling was therefore invaluable in elucidating the mechanism of triosephosphate isomerase (shown below) in which B is a protein-derived base.



Isotopes can be used in another way to measure the energy barrier heights for various steps in the catalytic mechanism as noted above for the reaction catalyzed by dihydrofolate reductase. For example, if a proton transfer is involved in the rate-limiting step, then substitution of that proton with one of the heavier isotopes of hydrogen (deuterium or tritium) will cause the step to proceed more slowly. These so-called kinetic isotope effect experiments in combination with steady-state rate measurements in the case of TIM allowed the elucidation of the rate constants for partitioning of the *cis*-enediol intermediate and construction of a detailed kinetic scheme as shown above for dihydrofolate reductase.

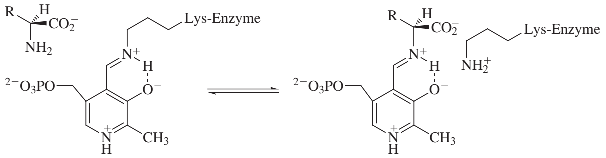
E. Aspartate Aminotransferase

Many enzymes employ exogenous molecules known as cofactors to assist in executing their chemistry. Sometimes these cofactors are covalently bound to the enzyme and sometimes not. Many types of cofactors are known, and here we will focus on a well-studied example called pyridoxal phosphate (PLP), which often participates in the metabolism of amino acids. PLP, derived from vitamin B_6, is a covalently bound cofactor; it is attached to lysine residues by means of a Schiff base or imine linkage as shown at right.



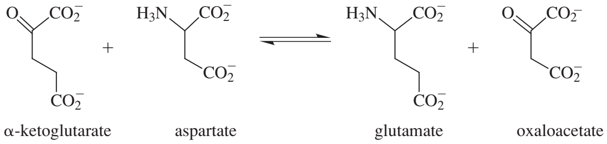
PLP enzyme-bound PLP

The substrates for most PLP-requiring processes are α-amino acids, and most of the processes take place at the α-carbon position, although some take place at the β- or γ-carbon. The enzymes which use PLP catalyze a wide range of reactions, including racemizations, decarboxylations, and amine transfers. In general, for all three of these classes of reactions at the α-carbon the substrate displaces the lysine and forms an aldimine intermediate with the PLP.



The now very acidic α-proton of the amino acid is abstracted by a basic amino acid residue (often the displaced lysine), with the pyridine ring of PLP acting as an electron sink. For the racemases, a proton is then delivered to the opposite face from the same or a different basic residue with the net result of inversion of configuration at the α-carbon. Attack of the active site lysine effects product release and regenerates the cofactor.

The structure of one PLP-utilizing transaminase, aspartate aminotransferase, is shown in Fig. 8. This enzyme catalyzes the reversible transamination reaction shown below.



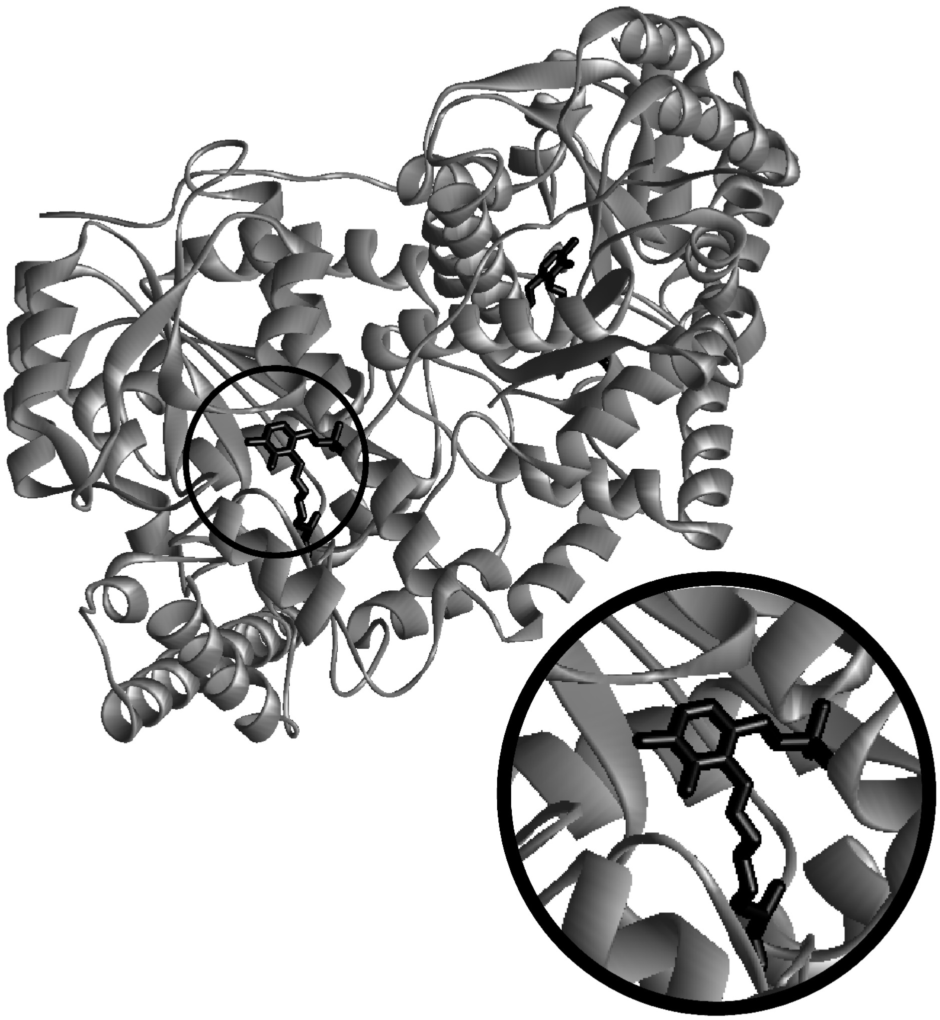
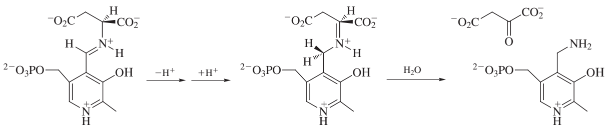


FIGURE 8 Structure of an aspartate aminotransferase. The protein is a homodimer, with one covalently bound pyridoxal phosphate (shown in black) in each of the two subunits. The expanded view shows the cofactor in greater detail. [Adapted from Rhee, S. *et al.* (1997). “Refinement and comparisons of the crystal structures of pig cytosolic aspartate aminotransferase and its complex with 2-methylaspartate,” *J. Biol. Chem.* **272,** 17293–17302.]

In the transamination reaction, formation of the aldimine intermediate between aspartate and PLP and its deprotonation proceeds as described above for the racemases. However, reprotonation occurs not at the same carbon as in the racemization mechanism but at a position adjacent to the PLP heterocycle.



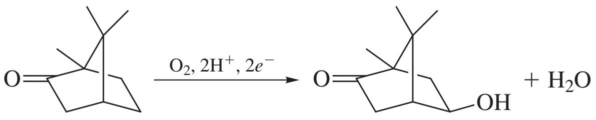
Hydrolysis releases the product oxaloacetate and generates a new form of the cofactor called pyridoxamine. The reverse reaction is then carried out on the other substrate, α-ketoglutarate, forming glutamate and regenerating the PLP cofactor.

Reactions at the β-position (for example, in threonine dehydatase) or the γ-position (in methionine-γ-lyase) also proceed by means of formation of an aldimine intermediate with the α-carbon of an α-amino acid. Such a survey of PLP-dependent enzymes illustrates the important point that one cofactor can be used for different kinds of transformations. The reactions described all go through a common aldimine intermediate, with the ultimate course of the reaction being controlled by the appropriate substrate specificity and positioning of amino acid side chains. This flexibility allows nature to expand its chemical repertoire with a relatively small set of cofactors.

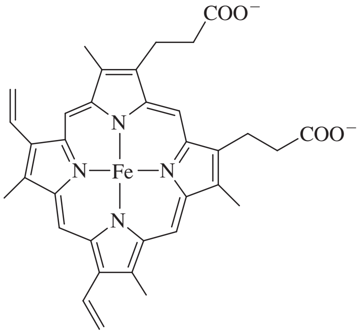
There are other organic cofactors such as thiamine pyrophosphate and biotin that participate in carbon–carbon bond formation and cleavage, cofactors that participate in reduction/oxidation, or redox, reactions such as nicotinamide and flavin moieties discussed in some of the earlier examples, and still others that are metal based such as vitamin B_{12} and porphyrin, which is our next topic.

F. Cytochrome P450

A different kind of cofactor from PLP is responsible for the chemistry of cytochrome P450 (Fig. 9), an enzyme which oxidizes hydrocarbons. It is known as a mixed-function oxidase, or monooxygenase, because one oxygen atom from molecular oxygen is incorporated into the product while the other goes on to form water. Cytochrome P450 in the liver, for example, oxidizes and detoxifies many kinds of substances that would otherwise be poisonous. One such well-studied reaction, the hydroxylation of camphor, is depicted below.



At the core of cytochrome P450 is an iron porphyrin, or heme, group, protoporphyrin IX, which is depicted below.



In the structure in Fig. 9, the iron porphyrin is shown in black.

Cytochrome P450 is a redox catalyst. The multiple available oxidation states allow the cofactor to accept and donate electrons during different stages of the catalytic cycle. Since the early 1970s, this enzyme and its relatives have been the subject of intense study by, among others, enzymologists, toxicologists, biophysical chemists, and inorganic chemists. The latter have tried to model the chemistry of cytochrome P450 with synthetic small molecules with the twin goals of mimicking its activity and understanding how the enzyme itself works. Working in parallel, biophysical chemists and enzymologists have performed many steady-state and pretransient kinetic studies such as the ones already discussed, which have contributed to a working model for the mechanism shown in Fig. 10.

Although this mechanism is in some senses more complicated than those that we have discussed, the same concepts apply. Starting at the top of the cycle, in the resting state of the enzyme the iron is in the +3 oxidation state and is bound by water. Substrate docks to its specific binding site and displaces water to start the catalytic cycle, and an electron is then introduced to reduce the iron to the +2 oxidation state. The dashed line is meant to indicate association of the substrate with the active site, not an actual bond to the iron. The requirement that substrate bind before reduction occurs is a control feature which prevents formation of very active and potentially damaging species in the absence of substrate. Oxygen then binds and accepts an electron from the iron, and introduction of another electron and two protons allows one atom of dioxygen to be released as water, which leaves behind a very active high valent (formally iron 5+) species. What follows is known as a radical rebound step. A hydrogen atom is removed from the substrate and transferred to the terminal oxygen atom, which produces a substrate radical. The radical recombines with the new hydroxo moiety to form the hydroxylated product, which

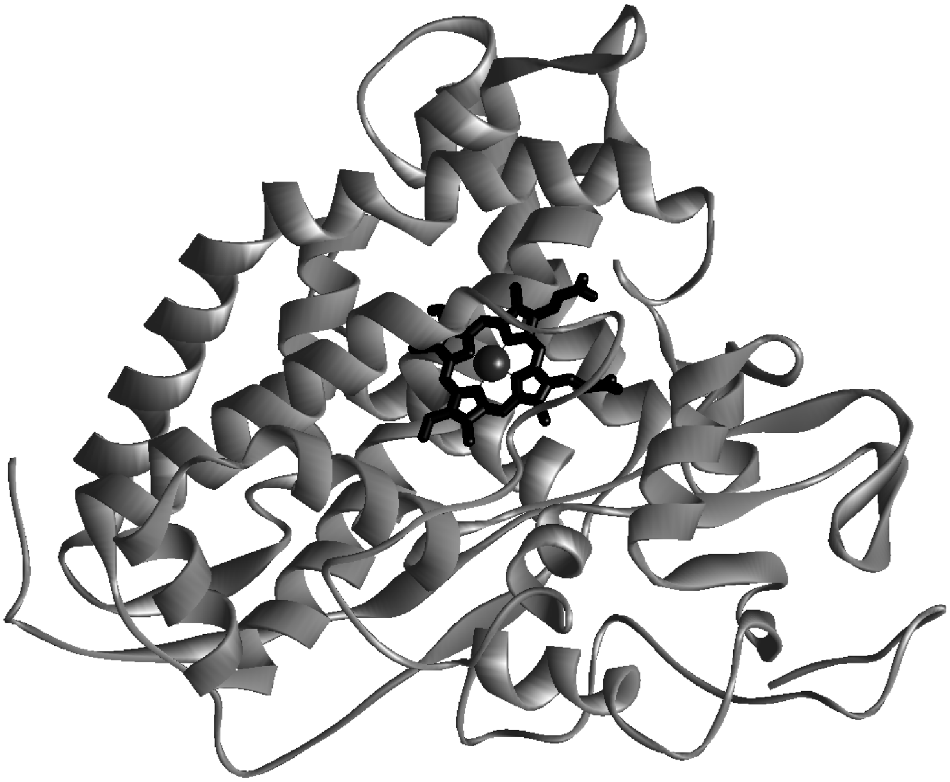


FIGURE 9 The cytochrome $P450_{cam}$ structure. The bound heme is depicted in black, and the iron atom at the center of the heme appears as a sphere. [Adapted from Poulos, T. L., Finzel, B. C., and Howard, A. J. (1986). "Crystal structure of substrate-free *Pseudomonas putida* cytochrome P450," *Biochemistry* **25,** 5314–5322.]

is then displaced by water; this completes the catalytic cycle.

One important line of investigation which has supported the radical rebound hypothesis is the use of radical clock substrate probes. These probes rearrange in a diagnostic way on a very rapid and calibrated time scale when a hydrocarbon radical is formed. In the case of P450, rearranged products have been isolated after oxidation and have been used as evidence of an intermediate substrate radical. In this way, even though the lifetime of the radical is too short for it to be observed directly, its character can be explored by the judicious choice of substrate analogues.

Mechanistic proposals are under constant scrutiny and revision, and aspects of the foregoing mechanism have been challenged. In particular, the possibility has been suggested that a species other than a high valent iron-oxo (likely a hydroperoxo species) may be the active oxidant for some substrates. Debates such as these are a great strength of the study of enzyme mechanisms. Given all the tools which have been developed in this field, and the wealth of interesting problems to which these tools can be applied, the study of enzyme mechanisms should be considered a vital and evolving process. The answer to the question of "how enzymes work" cannot be described fully in a single scheme.

IV. ORIGINS OF THE CATALYTIC EFFICIENCY OF ENZYMES

The source of the stereospecificity of enzyme-catalyzed reactions is clearly revealed by the fit of the substrate to the enzyme's active site that spatially then directs the stereochemical course of the chemical events. The speed of these reactions has been attributed to the lowering of the activation energy for the process by the greater affinity of the enzyme for the transition state than that for the substrate. Although this proposal is an adequate rationale, it is often a necessary thermodynamic statement that does not offer insights into how the activation barrier is actually lowered.

The preorganization of substrate and active site residues within a protein cavity converts an intermolecular process to intramolecular and may have both an enthalpic and an entropic advantage. The active site provides an environment in which the enzyme·substrate complex is

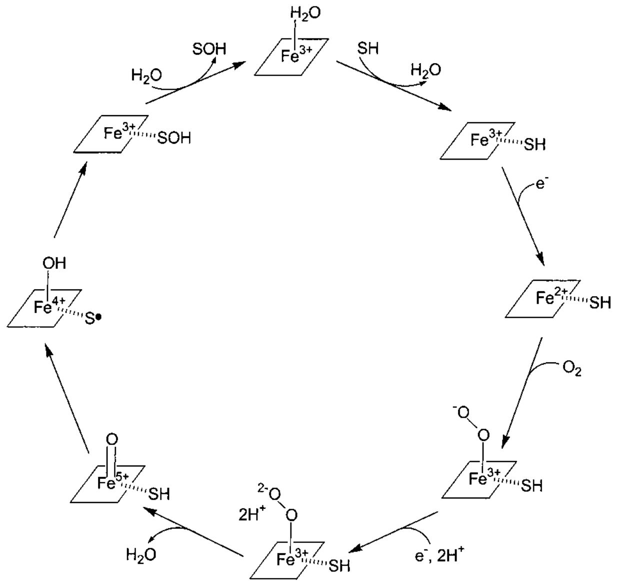


FIGURE 10 A currently accepted version of the catalytic cycle of cytochrome P450. The iron porphyrin is drawn as a parallelogram, the substrate is designated as *S*H and the product as *S*OH. See the text for a description. [Adapted from Mueller, E. J., Loida, P. J., and Sligar, S. G. (1995). Twenty-five years of $P450_{cam}$ research. *In* "Cytochrome P450: Structure, Mechanism, and Biochemistry" (P. R. Ortiz de Montellano, ed.), 2nd ed., pp. 83–124, Plenum, New York.]

populated with cofactors that are poised for reaction. These structures, or NACs (near attack conformers), are similar in structure to the transition state so that only slight changes in bond distances and angles within the structures through the normal dynamic motions of the protein are sufficient to trigger the crossing of the reaction barrier. The enzyme's active site is also preorganized in the sense that the locus of general acids/bases, nucleophiles, solvents, dipoles, hydrogen bonds, and so forth are fixed by the NAC to interact with the transition state. Molecular dynamic calculations sampling several enzyme classes suggest that the affinity of these enzymes for their transition states is little changed from that for the substrate. The enzyme-catalyzed reaction also benefits in many cases due to the nonaqueous interaction of the active site cavity, which can often accelerate, by large factors, the reaction over that in aqueous media.

For DHFR in particular, molecular dynamics calculations, NMR measurements of solution structure, and kinetics measurements of mutant forms of the enzyme appear to support the importance of dynamic motions of the protein fold to trigger the reaction of an enzyme–substrate NAC. The mutations in question (for example Gly120 in Fig. 7) are well removed from the active site and underscore the role of the entire protein fold. The contribution of dynamic motions to the overall catalytic rate remains to be elucidated for the majority of enzymes. Their existence may explain why more rigid molecules such as imprinted polymers and catalytic antibodies do not generally exhibit the large rate accelerations noted with enzymes despite the fact that they too have converted an intermolecular process to an intramolecular process.

SEE ALSO THE FOLLOWING ARTICLES

Bioconjugate Chemistry • Bioenergetics • Bioinorganic Chemistry • Bioreactors • Fiber-Optic Chemical Sensors • Gene Expression, Regulation of • Lipoprotein/Cholesterol Metabolism • Translation of RNA to Protein • Vitamins and Coenzymes

BIBLIOGRAPHY

Bugg, T. (1997). "An Introduction to Enzyme and Coenzyme Chemistry," Blackwell Sci., London.

Fersht, A. (1999). "Structure and Mechanism in Protein Science," Freeman, New York.

Hammes, G. G. (1982). "Enzyme Catalysis and Regulation," Academic Press, New York.

Ortiz de Montellano, P. R., ed. (1995). "Cytochrome P450: Structure, Mechanism, and Biochemistry," 2nd ed., Plenum, New York.

Page, M. I. (1984). "The Chemistry of Enzyme Action," Elsevier, New York.

Price, N. C., and Stephens, L. (1989). "Fundamentals of Enzymology," Oxford Univ. Press, New York.

Food Colors

Pericles Markakis
Michigan State University

I. Introduction
II. Natural Food Pigments
III. Food Browning
IV. Color Additives in Foods

GLOSSARY

Anthocyanins Red, blue, and violet water-soluble plant pigments of a phenolic nature.

Browning, food Darkening of foods as a result of enzymatic or nonenzymatic reactions.

Caramel Brown coloring matter made by heating sugars dry or in solution.

Carotenes Chiefly orange-yellow plant and animal pigments; some are provitamins A.

Certification, color Submission of a sample of a listed color additive to the Food and Drug Administration and, after chemical analysis, issuance of a certificate permitting marketing of the batch from which the sample was taken; certain color additives are exempt from certification.

Chlorophyll Green pigment of plants; chemically it is related to the red pigment of blood.

Colorant Substance that colors or modifies the color of another substance.

Excipient Inert substance used as a diluent or vehicle of a colorant.

Heme Color-furnishing portion of the red pigment molecule of blood and meat.

Lakes, color Water-insoluble pigments prepared by precipitating soluble dyes on an insoluble substratum, alumina in the case of food lakes.

Listed color(ant)s Color additives that have been sufficiently evaluated to convince the Food and Drug Administration of their safety for the application intended.

FOOD COLORS are both the sensations evoked when light reflected from foods stimulates the retina of the eye and the particular food components involved in the process. These components, also known as food colorants, may be present in foods naturally, or formed during food processing, or intentionally added to foods, or all of these. This article deals with all groups of food colorants.

I. INTRODUCTION

Color is important for identifying foods, judging their quality, and eliciting aesthetic pleasure in our encounters with them. Because color is usually the first food attribute to strike the senses, its significance in food marketing is obvious ("eating" with the eyes). Thus, all food providers (growers, grocers, homemakers, chefs, and industrial food processors) do their best to present a food with an attractive

color. In certain instances, the original color of the food must be preserved, as is the case with most fruits and vegetables. In other instances, culinary art is required to create new, pleasing colors, as when turkey is roasted, bread is baked, or potato chips are fried. In still other instances, colors (colorants) are added to foods, as is done with many beverages and candies.

The coloring matter of foods is discussed under three headings: natural food colors, food browning, and food color additives.

II. NATURAL FOOD PIGMENTS

Approximately 1500 colored compounds, also known as natural food pigments, have been isolated from foodstuffs. On the basis of their chemical structure, these food pigments can be grouped in the following six classes: heme pigments, chlorophylls, carotenoids, flavonoids, betalains, and miscellaneous pigments.

A. Heme Pigments

Heme (from the Greek for blood) is the basic chemical structure (Fig. 1) responsible for the red color of two important animal pigments: hemoglobin, the red pigment of blood, and myoglobin, the red pigment of muscles. Practically all the red color of red meat is due to myoglobin, since the hemoglobin is removed with the bleeding of the slaughtered animal. Other colored muscle compounds (cytochromes, vitamin B_{12}, flavoproteins) do not contribute significantly to the color of red meat.

Myoglobin is a protein that facilitates the transfer of oxygen in muscles. It was the first protein to be fully elucidated with regard to the three-dimensional arrangement of its atoms. Hemoglobin, the oxygen-carrying pigment of blood, is composed of four heme groups attached to four polypeptide chains.

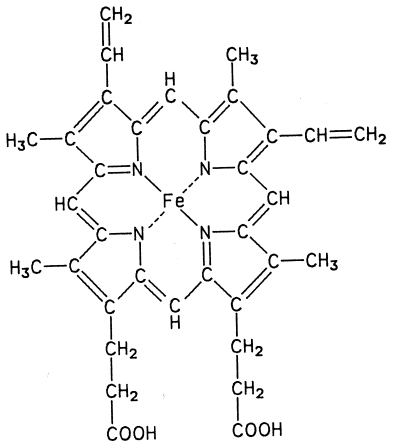


FIGURE 1 Structure of heme.

The myoglobin in meat is subject to chemical and color changes. Freshly cut meat looks purplish. On exposure to air, the surface of the meat acquires a more pleasing red hue (blooming of the cut). The color change is due to the oxygenation of myoglobin (an oxygen molecule is attached to the heme group in a fashion parallel to the oxygenation of hemoglobin). The oxygenated myoglobin is called oxymyoglobin. When meat is packed in plastic film, the oxygen permeability of the film should be sufficient to keep the myoglobin oxygenated. In both myoglobin and oxymyoglobin the heme iron is in the Fe^{2+} form. In the presence of oxygen, myoglobin is eventually oxidized to brown metmyoglobin, in which the heme iron is in the Fe^{3+} form. Both the oxygenation and oxidation processes are reversible. Severe oxidative deterioration may result in the formation of green pigments (sulfmyoglobin, cholemyoglobin).

When meat is cooked, the protein moiety (globin) of myoglobin is denatured and the heme is converted chiefly to nicotinamide hemichrome, the entire pigment acquiring a brown hue. These changes are irreversible. Heated meat is also subject to the browning reactions discussed in Section III. A simplified scheme of the red-pigment changes in fresh and heated meat is shown in Fig. 2.

In cured meats, in which nitrite is used, many reactions occur, some of which lead to color changes. Among the established reactions are the following: (1) the nitrite salt is converted to nitric oxide (NO), nitrate, and water; (2) the NO replaces the H_2O attached to the iron of heme and forms nitrosyl myoglobin, which is reddish; (3) on heating, the nitrosyl myoglobin is transformed to nitrosyl hemochrome, which has the familiar pink color of cured meats; and (4) any metmyoglobin present in the cured meat is similarly nitrosylated, reduced, and finally converted to nitrosyl hemochrome.

B. Chlorophylls

Several chlorophylls have been described. Two of them, chlorophyll *a* and chlorophyll *b*, are of particular interest in food coloration because they are common in green plant tissues, in which they are present in the approximate ratio 3 : 1, respectively. Their structures resemble that of heme since they are all derivatives of tetrapyrrole. An important difference is that the central metal atom is iron in heme and magnesium in the chlorophylls. Another difference is that the pyrrole unit IV in the chlorophylls is hydrogenated. In addition, the chlorophylls contain a 20-carbon hydrophobic "tail," the phytyl group (Fig. 3).

The chlorophylls are located in special cellular bodies, the chloroplasts, where they function as photosynthetic

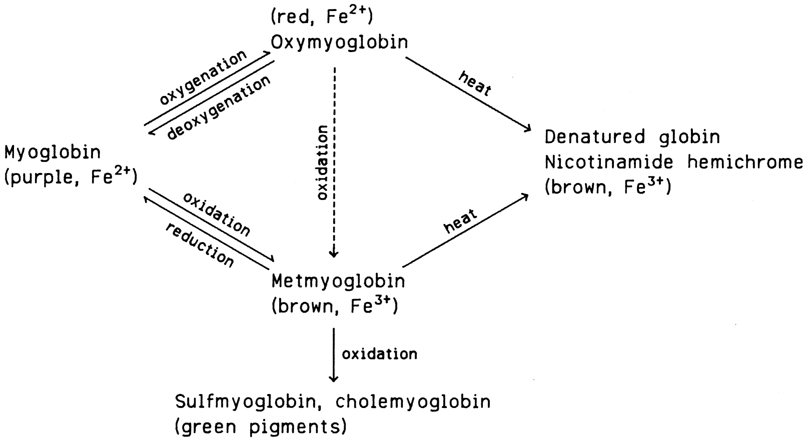


FIGURE 2 Pigment changes in fresh and heated red meat.

agents. As food pigments, chlorophylls impart their green color to many leafy (spinach, lettuce, etc.) and nonleafy (green beans and peas, asparagus, etc.) vegetables and to unripe fruits. They are not very stable pigments, however. Ethylene, a gaseous plant hormone, destroys chlorophylls, and it is occasionally used to degreen fruits. The acids naturally present, formed, or added to plant tissues during food processing convert the bright green chlorophylls to dull olive brown pheophytins by replacing the magnesium of the molecule with hydrogen. Unfortunately, no fail-safe procedure has been proposed for preventing this discoloration in heated and stored green vegetables. Freezing storage is an effective method of preserving the green color of vegetables.

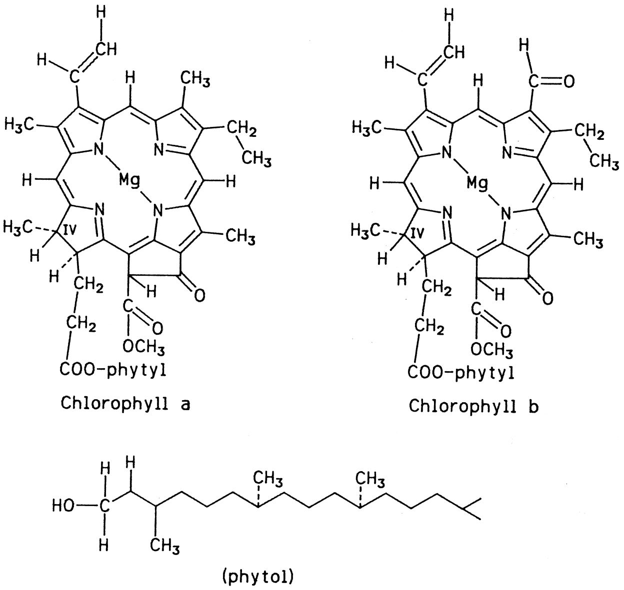


FIGURE 3 Structure of chlorophylls *a* and *b*. [From Aronoff, S. (1966). *In* "The Chlorophylls" (L. P. Vernon and G. R. Seeley, eds.), Academic Press, New York.]

C. Carotenoids

Many of the yellow, orange, and red colors of plants and animals are due to carotenoids, pigments similar to those of carrots. The basic structure of carotenoids is a chain of eight isoprenoid units. Certain isoprenoid derivatives with shorter chains (e.g., vitamin A) are also considered carotenoids. Most of the structural differences among carotenoids exist at the ends of the chain. Some carotenoids are hydrocarbons and are known as carotenes, while others contain oxygen and are called xanthophylls. The structures of several carotenoids, along with the foods or tissues in which they are present, are shown in Table I.

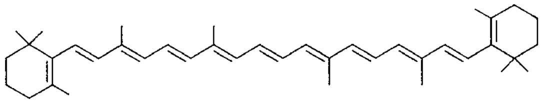
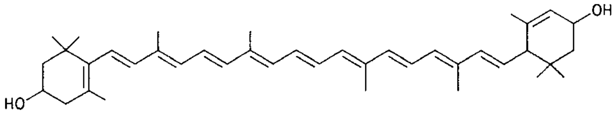
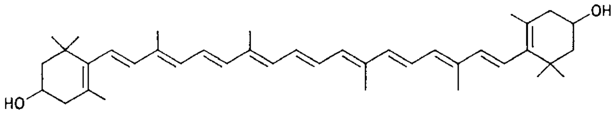
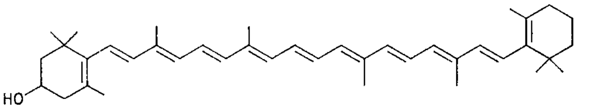
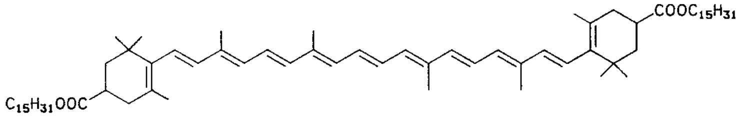
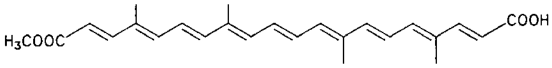
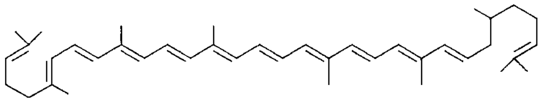
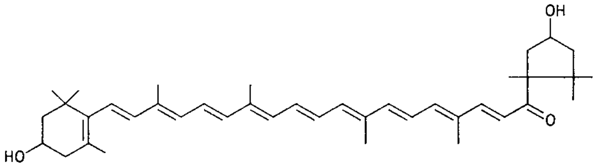
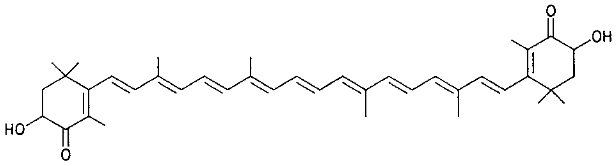
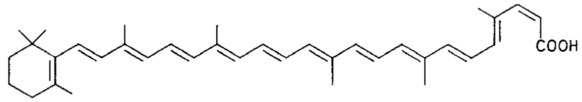
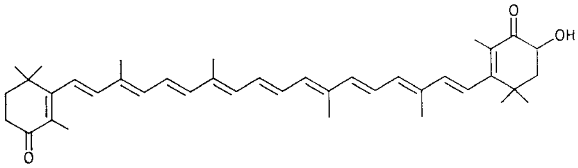
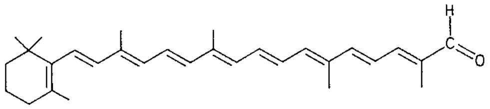
Because of the numerous double bonds in the carotenoid molecule, a large number of cistrans isomers are theoretically possible. The carotenoids of foods, however, are usually in the all-trans form (Table I). Trans to cis transformation is possible and is accelerated by heat, light, and acidity.

Carotenoids occur free or as esters of fatty acids or as complexes with proteins and carbohydrates; for example, in paprika, capsanthin is esterified with lauric acid. In live lobster, astaxanthin is complexed with protein; the astaxanthin–protein complex is blue-gray, the color of live lobster, but on heating, the complex is broken and the freed astaxanthin imparts its red color to the cooked lobster.

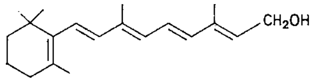
Carotenoids are present in a large variety of foods, from yeast and mushrooms, to fruits and vegetables, to eggs, to fats and oils, to fish and shellfish. As fat-soluble substances, carotenoids tend to concentrate in tissues or products rich in lipids, such as egg yolk and skin fat, vegetable oils, and fish oils.

Plants and microorganisms synthesize their own carotenoids, while animals appear to obtain theirs from primary producers. In the development of many fruits (e.g., citrus fruits, apricots, tomatoes) ripening is associated with the accumulation of carotenoids and the

TABLE I Types of Carotenoids and Their Natural Sources

Structure	Name and source
	β-Carotene (carrot, egg, orange, chicken fat)
	Xanthophyll (vegetables, egg, chicken fat)
	Zeaxanthin (yellow corn, egg, liver)
	Cryptoxanthin (egg, yellow corn, orange)
	Physalien (asparagus, berries)
	Bixin (annatto seeds)
	Lycopene (tomato, pink grapefuit, palm oil)
	Capsanthin (paprika)
	Astaxanthin (lobster, shrimp, salmon)
	Torularhodin (*Rhodotorula* yeast)
	Canthaxanthin (mushrooms)
	β-Apo-8′-carotenal (spinach, orange)

disappearance of chlorophyll. The intensity of the yellow color of certain animal products, such as egg yolk and milk fat or butter, depends on the carotenoid content of the feed the animals ingest. In view of this dependency, the seasonal variation in the color of these products is understandable. A nutritionally important interconversion of carotenoids is the formation of retinol (vitamin A) from



β-carotene and other carotenoids possessing a β-ionone ring and known as provitamins A.

The stability of carotenoids in foods varies greatly, from severe loss to actual gain in carotenoid content during storage. Carotenoid losses amounting to 20 or 30% have been observed in dehydrated vegetables (e.g., carrots, sweet potatoes) stored in air. These losses are minimized when the dry product is stored in vacuum or inert gas (e.g., nitrogen), at low temperatures, and protected from light. The main degradative reaction of carotenoids is oxidation. Oxygen may act either directly on the double bonds or through the hydroperoxides formed during lipid autoxidation. Hydroperoxides formed during enzymatic lipid oxidation can also bleach carotenoids by a coupled lipid–carotenoid oxidation mechanism. On the other hand, certain vegetables, such as squash and sweet potatoes, in which carotenoid biosynthesis continues after harvesting, may manifest an increase in carotenoid content during storage.

D. Flavonoid Pigments

Hundreds of flavone-like pigments are widely distributed among plants. On the basis of their chemical structure, these pigments are grouped in several classes, the most important of which are listed in Table II. The basic structure of all these compounds comprises two benzene rings, A and B, connected by a heterocycle. The classification of flavonoids is based on the nature of the heterocycle (which is open in one class).

Most of these pigments are yellow (Latin, *flavus*). One important exception is the anthocyanins, which display a great variety of red and blue hues. Because of the strong visual impact of anthocyanins on the marketing of fruits and vegetables, these pigments will be discussed in greater detail than other flavonoids.

1. Anthocyanins

The name of these pigments was originally coined to designate the blue (*kyanos*) pigments of flowers (*anthos*). It is now known that not only the blue color, but also the purple, violet, magenta, and most of the red hues of flowers, fruits, leaves, stems, and roots are attributable to pigments chemically similar to the original "flower blues." Two exceptions are notable: tomatoes owe their red color to lycopene and red beets owe theirs to betanin, pigments not belonging to the anthocyanin group.

Anthocyanins are glycosides of anthocyanidins, the latter being polyhydroxyl and methoxyl derivatives of flavylium. The arrangement of the hydroxyl and methoxyl groups around the flavylium ion in six anthocyanidins common in foods is shown in Fig. 4.

There are at least 10 more anthocyanidins in nature, practically always appearing as glycosides. The number of anthocyanins far exceeds that of anthocyanidins, since monosaccharides, disaccharides, and at times trisaccharides glycosylate the anthocyanidins at various positions (always at 3, occasionally at 5, and seldom at other positions). Eventual acylation with *p*-coumaric, caffeic, and ferulic acids increases the number of natural anthocyanins. An example of acylated anthocyanin is the dark purple eggplant pigment delphinidin, 3-[4-(p-coumaroyl)-L-rhamnosyl-(1 → 6)-D-glycosido] 5-D-glucoside.

The color of anthocyanins is influenced not only by structural features (hydroxylation, methoxylation, glycosylation, acylation), but also by the pH of the solution in which they are present, copigmentation, metal complexation and self-association.

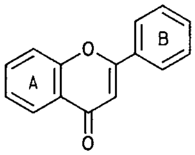
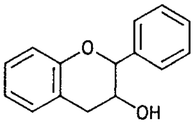
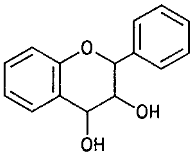
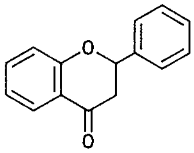
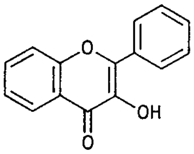
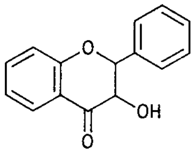
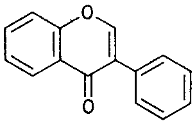
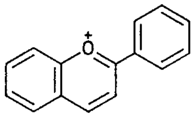
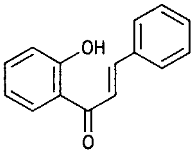
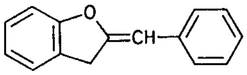
The pH affects both the color and the structure of anthocyanins. In very acidic solution, anthocyanins are red, but as the pH rises the redness diminishes. In freshly prepared alkaline or neutral solution, anthocyanins are blue or violet, but (with the exception of certain multiacylated anthocyanins) they fade within hours or minutes.

In acidic solution four molecular species of anthocyanins exist in equilibrium: a bluish quinoidal (or quinonoidal) base A, a red flavylium cation AH^+, a colorless carbinol pseudo-base B, and a colorless or yellowish chalcone C (Fig. 5).

At very low pH (below 1), the red cation AH^+ dominates, but as the pH rises to 4 or 5, the concentration of the colorless form B increases rapidly at the expense of AH^+, while forms A and C remain scarce. In neutral and alkaline solutions, the concentration of base A rises and its phenolic hydroxyls ionize, yielding unstable blue or violet quinoidal anions A^- (Fig. 6).

Although it is true that the reaction of most plant tissues pigmented with anthocyanins (fruits, flowers, leaves) is slightly acidic, pH alone cannot explain the vivid colors encountered in these tissues. One mechanism leading to the enhancement and stability of anthocyanin coloration is copigmentation, that is, the association of anthocyanins with other organic substances (copigments). This association results in complexes that absorb more

TABLE II Major Classes of Flavonoids

Class	Structure	Example (source)
Flavones		Apigenin (chamomile)
Flavan-3-ols (catechins)		(−)-Epicatechin (cocoa)
Flavan-3, 4-diols		Leucocyanidin (peanut)
Flavanones		Naringenin, hesperidin (citrus fruits)
Flavonols		Quercetin (apples, grapes)
Flavanonols		Taxifolin (*Prunus*)
Isoflavones		Genistein (soybeans)
Anthocyanidins		Pelargonidin, cyanidin, delphinidin (berries, red apples, red grapes)
Chalcones		Butein (*Butea*)
Aurones		Aureusidin (*Oxalis*)

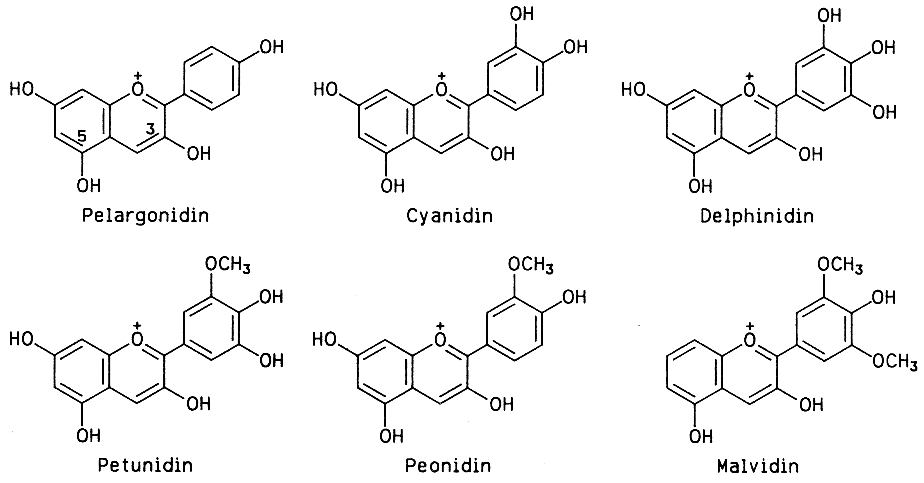


FIGURE 4 Six anthocyanidins common in foods. The electric charge shown at position 1 is delocalized over the entire structure by resonance.

visible light (they are brighter) and light of lower frequency (they look bluer—the bathochromic effect) than the free anthocyanins at tissue pH. Most of these copigments are flavonoids, although compounds belonging to other groups (e.g., alkaloids, amino acids, nucleotides) can function similarly. A stacked molecular complex between an acylated anthocyanin and a copigment (flavocommelin) is shown in Fig. 7.

Self-association is the binding of anthocyanin molecules to one another. It has been observed that the complexes absorb more light than the sum of the single molecules. This explains why a 100-fold increase in the concentration of cyanidin 3,5-glucoside results in a 300-fold rise in absorbance.

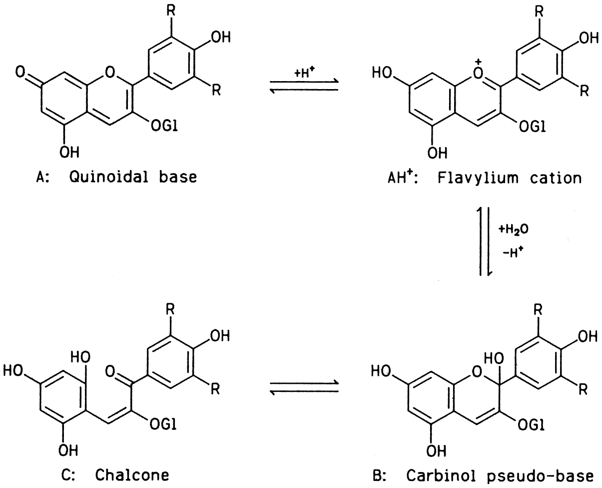


FIGURE 5 Four anthocyanin structures present in aqueous acidic solutions: R is usually H, OH, or OCH_3. Gl is glycosyl. [Adapted from Brouillard, R. (1982). *In* "Anthocyanins as Food Colors" (P. Markakis, ed.), Academic Press, New York.]

Certain anthocyanins form complexes with metals (e.g., iron, aluminum, magnesium), and the result is an augmentation of the anthocyanin color. At times the complexes involve an anthocyanin, a copigment, and a metal.

A large number of the anthocyanins present in fruits and vegetables have been identified. It is not unusual for a plant tissue to contain several anthocyanins (17 in certain grape varieties), all genetically controlled. Table III shows the anthocyanidin moieties of anthocyanins in common fruits and vegetables.

Generally, the attractive color of anthocyanin-pigmented foods is not very stable. Canning of red cherries or berries results in products with considerable bleaching. Strawberry preserves lose one-half of their anthocyanin content after a few weeks on the shelf, although the browning reaction may mask the loss. And red grape juice is subject to extensive color deterioration during storage.

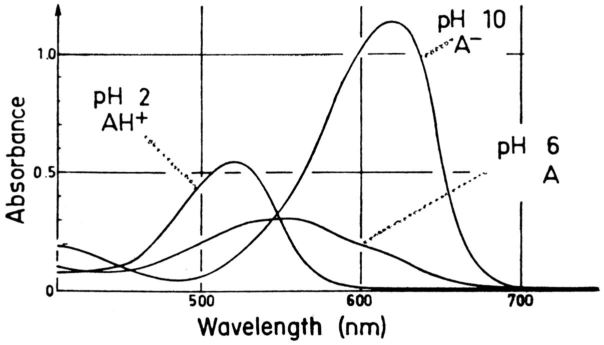


FIGURE 6 Absorption spectra recorded immediately after dissolving an anthocyanin (malvin chloride) in buffers of pH 2, 6, and 10. The absorption peaks at pH 6 and 10 disappeared within 1 to 3 hr. (Adapted from Brouillard, R. (1982). *In* "Anthocyanins as Food Colors" (P. Markakis, ed.), Academic Press, New York.]

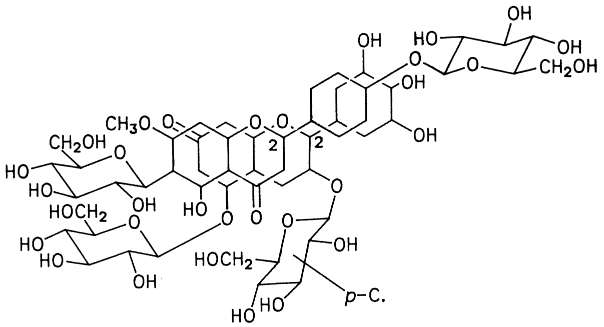


FIGURE 7 Stacked molecular complex of awobanin and flavocommelin; *p*-C. denotes *p*-coumaroyl. [From Osawa, Y. (1982). *In* "Anthocyanins as Food Colors" (P. Markakis, ed.), Academic Press, New York.]

Exposure to high temperatures and contact with the oxygen of the air appear to be two factors affecting anthocyanin stability most adversely. Ascorbic acid accelerates the destruction of anthocyanins, and so does light. Certain oxidizing enzymes, such as phenol oxidase, and a hydrolyzing enzyme known as anthocyanase may contribute to the degradation of anthocyanin pigments. Oxidizing enzymes act on the anthocyanidin moiety, while anthocyanase splits off the sugar residue(s); the freed anthocyanidin is very unstable and loses its color spontaneously. Sulfur dioxide, which is used for the preservation of some fruit products (pulps, musts), bleaches anthocyanin pigments, but on heating of the fruit prduct in vacuum the SO_2 is removed and the anthocyanin color reappears. Large concentrations of SO_2, combined with lime, decolorize anthocyanins irreversibly and are used in the preparation of maraschino cherries. Anthocyanins act as anodic and cathodic depolarizers and thereby accelerate the internal corrosion of tin cans. It is therefore necessary to pack anthocyanin-colored products in cans lined with special enamel. In aging red wines anthocyanins condense with other flavonoids and form polymeric (MW $\leq$ 3000) redbrown pigments (Fig. 8). On continued polymerization these pigments become insoluble and form sediments in bottled red wines.

Anthocyanins possessing more than one acyl group show extraordinary color stability over a wide pH range. One of them, peonidin-3-(dicaffeyl sophoroside) 5-glucoside, isolated from 'Heavenly Blue' morning glory flowers (*Ipomoea tricolor*), has been shown to "produce a wide range of stable colors in foods and beverages which have a pH range of 2.0 to about 8.0." United States patent 4,172,902 covers its use as a colorant in foods.

2. Other Flavonoids

Among flavonoids other than anthocyanins, the catechins, flavonols, and leucoanthocyanidins have the widest distribution in foodstuffs, while flavonone glycosides are of special interest in citrus fruits.

Catechins, or flavan-3-ols, are present mainly in woody tissues. Among common foods, tea leaves contain at least six catechins representing about 25% of the dry weight of tea leaves. Tea catechins are excellent substrates for the catechol oxidase that is present in tea leaves and participates in the conversion of green tea to black tea. The reddish brown color of tea brew is due to a mixture of

TABLE III Anthocyanidins Present as Anthocyanins in Fruits and Vegetables

Fruit or vegetable	Anthocyanidin
Apple (*Malus pumila*)	Cyanidin
Blackberry (*Rubus fructicosus*)	Cyanidin
Black currant (*Ribes nigrum*)	Cyanidin. delphinidin
Blueberry (lowbush, *Vaccinium angustifolium;* highbush, *V. corymbosum*)	Delphinidin, petunidin, malvidin, peonidin, cyanidin
Cherry (sour, 'Montmorency,' *Prunus cerasus;* sweet, 'Bing,' *P. avium*)	Cyanidin, peonidin
Cranberry (*Vacinnium macrocarpon*)	Cyanidin, peonidin
Elderberry (*Sambucus nigra*)	Cyanidin
Fig (*Ficus carica*)	Cyanidin
Gooseberry (*Ribes grossularia*)	Cyanidin
Grape (red European. *Vitis vinifera*)	Malvidin, peonidin, delphinidin, cyanidin, petunidin, pelargonidin
Grape ('Concord,' *Vitis labrusca*)	Cyanidin, delphinidin, peonidin, malvidin, petunidin
Mango (*Mangifera indica*)	Peonidin
Mulberry (*Morus nigra*)	Cyanidin
Olive (*Olea europea*)	Cyanidin
Orange ('Ruby,' *Citrus sinesis*)	Cyanidin. delphinidin
Passion fruit (*Passiflora edulis*)	Delphinidin
Peach (*Prunus persica*)	Cyanidin
Pear (*Pyrus communis*)	Cyanidin
Plum (*Prunus domestica*)	Cyanidin, peonidin
Pomegranate (*Punica granatum*)	Delphinidin, cyanidin
Raspberry (*Rubus ideaus*)	Cyanidin
Strawberry (*Fragaria chiloensis* and *F. virginiaca*)	Pelargonidin, little cyanidin
Beans (red, black; *Phaseolus vulgaris*)	Pelargonidin, cyanidin, delphinidin
Cabbage (*red, Brassica oleracea*)	Cyanidin
Corn (red, *Zea mays*)	Cyanidin, pelargonidin
Eggplant (*Solanum melongena*)	Delphinidin
Onion (*Alium cepa*)	Cyanidin, peonidin
Potato (*Solanum tuberosum*)	Pelargonidin, cyanidin, delphinidin, petunidin
Radish (*Raphanus sativus*)	Pelargonidin, cyanidin

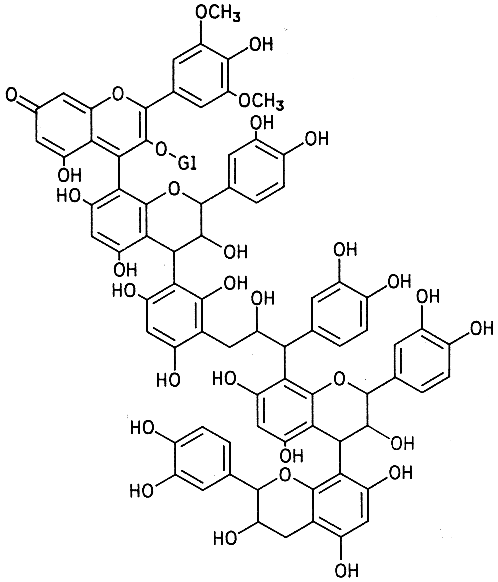


FIGURE 8 Proposed structure for a polymeric red-brown pigment in aging red wine. [From Somers, T. C. (1971). *Phytochemistry* **10,** 2184.]

pigments known as theaflavins and thearubigins. The structure of one of them is shown in Fig. 9.

Flavonols, like anthocyanidins, exist almost exclusively as glycosides. Three common flavonols are kaempferol, quercetin, and myricetin, resembling pelargonidin, cyanidin, and delphinidin, respectively, in the hydroxylation pattern of the B ring. Flavonol glycosides impart weak yellow hues to apples, apricots, cherries, cranberries, grapes, onions, plums, potatoes, strawberries, tea, tomatoes, and other commodities.

Leucoanthocyanidins are compounds of the general formula **1** shown in Fig. 10. They have no color of their own, but in acidic environments and at elevated temperatures they are converted to colored anthocyanidins (**2**). This reaction is in competition with the condensation to a dimeric leucoanthocyanidin (**3**). Low temperature favors the formation of the dimeric compound, which can polymerize to yield products with pronounced tanning properties.

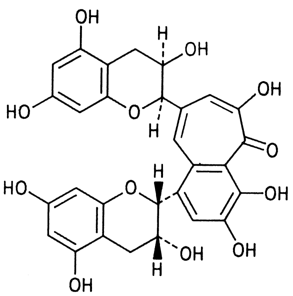


FIGURE 9 Structure of theaflavin.

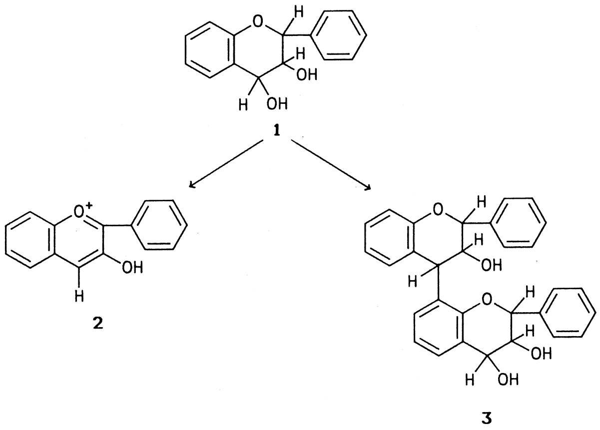


FIGURE 10 Basic structures of leucoanthocyanidins (**1**), anthocyanidins (**2**), and dimeric leucoanthocyanidins (**3**).

The most common leucoanthocyanidins are leucopelargonidin, leucocyanidin, and leucodelphinidin, which are converted to the corresponding anthocyanidins. This conversion results in the undesirable "pinking" of certain products such as canned pears, canned banana puree, processed brussels sprouts, and beer. On the other hand, polymerization to tannins leads to astringency and the formation of haze in beer (insolubilization of beer proteins).

E. Betalains

Betalain is a relatively new term used to describe a class of water-soluble plant pigments exemplified by the red-violet betacyanins and yellow betaxanthins. (In a parallel fashion, flavonoids comprise the red-blue anthocyanins and the typical yellow flavonoids that some authors call anthoxanthins.) Betalains owe their name to the red beet (*Beta vulgaris*), from which they were originally extracted, and they are not as widely distributed as flavonoids. Other foods containing betalains include chard, pokeberries, and Indian cactus fruits. The major red pigment of red beets is betanin, and their major yellow pigment is vulgaxanthin (Fig. 11).

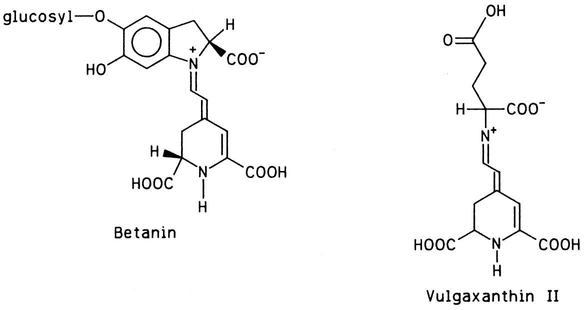


FIGURE 11 Two major pigments of red beets.

Betalains are stable in the pH range 3.5–7.0, which is the pH range of most foods, but they are sensitive to heat, oxidation, and light.

F. Miscellaneous Natural Food Colors

There are several hundred additional natural pigments that are not as widely represented in foods as the previously discussed coloring substances. Among them are the quinones and xanthones, which are yellow pigments. An example of a quinone is juglone, which is present in walnuts and pecans. Mangiferin, a representative of xanthones, is found in mangoes. Tannins include two types of pale yellow to light brown compounds, characterized by their property to convert animal hides to leather. One type consists of condensed tannins, to which reference was made in relation to the leucoanthocyanidins, and the other type comprises hydrolyzable tannins, which are esters of a sugar, usually glucose, with gallic acid, ellagic acid, or both. Corilagin is an example of a gallotannin, in which glucose is esterified with three gallic acid molecules. A yellow pigment that has attracted much attention because of its toxicity to humans and nonruminant animals is gossypol. It is present in cottonseeds, which are used as animal feed and have been considered a potential source of protein for human use. Several biologically very important food constituents are colored, such as phytochrome (yellow), vitamin B_2 (riboflavin, orange-yellow), and vitamin B_{12} (red), although their contribution to food coloration is negligible.

III. FOOD BROWNING

Foods may develop a variety of brown colors, from yellow-brown to red-brown to black-brown, during handling, processing, and storage. These colors are desirable in certain foods (e.g., coffee, beer, bread, maple syrup). In other foods, such as most dehydrated fruits and vegetables, dried eggs, and canned or dried milk, browning is detrimental. Even when desirable, browning should not be excessive, as in potato chips, french fries, and apple juice. Numerous reactions lead to browning in foods. Some of these may also generate flavors and/or alter the nutritional properties of foods. Conventionally, browning is discussed as enzymatic and nonenzymatic browning.

A. Enzymatic Browning

Several enzymes may initiate reactions that eventually produce brown colors in foods. For example, the action of ascorbate oxidase on ascorbic acid or of lipoxidase on lipids leads to carbonyl products that may either polymerize or react with amino compounds and form brown products. Phenolase (or phenol oxidase), however, is the principal browning enzyme. This enzyme oxidizes *o*-diphenols to *o*-quinones, which, by nonenzymatic processes, are ultimately converted to brown polymers known as melanins. Melanins are formed in both animal and plant tissues. A typical substrate of phenolase in animals is tyrosine. This amino acid is converted to melanin by a series of reactions, some of which are shown in Fig. 12.

In dark hair, skin, eyes, and other animal tissues, melanin is attached to proteins. Tyrosine is also a phenolase substrate in plant tissues (e.g., potatoes), but *o*-diphenols and polyphenols are by far the most common substrates of enzymatic browning in foods of plant origin. The following phenolic compounds have been associated with enzymatic browning in some foods: chlorogenic acid, caffeic acid, and catechin in apples, apricots, peaches, and pears; 3,4-dihydroxyphenylethylamine in bananas; (−)-epicatechin, (+)-catechin, (+)-gallocatechin, and (−)-epigallocatechin galate in tea leaves and cocoa beans; catechins in grapes; and tyrosine and chlorogenic acid in potatoes. The structures of four of these phenolics are shown in Fig. 13.

Many fresh fruits and vegetables brown slowly as they senesce. The enzymatic browning of these commodities is more rapid when they are subjected to processing, such as the pressing of apples in making cider or the peeling and cutting of potatoes in preparing potato products. Since enzyme, substrate, and oxygen must all be present for the development of this type of browning, elimination of any one of the three agents will prevent the browning. Heat inactivation of the enzyme, the exclusion of oxygen (by keeping the commodity under water or packaging it under vacuum or inert gas), and the selection of varieties poor in substrate content or enzyme activity are ways of preventing this discoloration. Also, storage at low temperature and the addition of sulfur dioxide, ascorbic acid, citric acid, sodium chloride, or combinations of these compounds will inhibit browning.

B. Nonenzymatic Browning

A number of chemical processes not involving enzymes may result in food browning. Briefly discussed here are the Maillard reaction, caramelization, ascorbic acid browning, and metalpolyphenol browning.

1. The Maillard Reaction (Maillard Browning)

This reaction is actually a series of reactions occurring from the first encounter of a carbonyl compound with an amine compound to the formation of brown pigments. It is also known as the carbonyl–amine reaction, and its brown

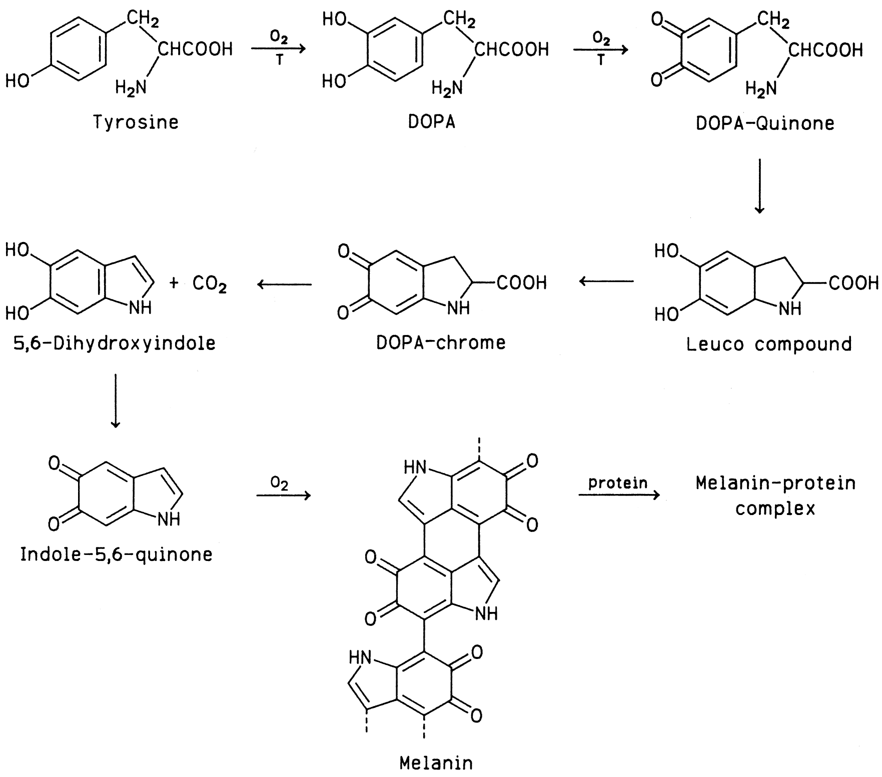


FIGURE 12 Conversion of tyrosine to melanin, catalyzed in part by tyrosinase (T). DOPA, Dihydroxyphenylalanine. Only part of melanin is shown.

products are often called melanoidins, indicating their visual similarity to the melanins of enzymatic browning. The most common carbonyl compounds of foods involved in the Maillard reaction are reducing sugars, and the most common amine compounds are amino acids.

The intermediate reactions and their relative velocities vary with the type of initial reactants and the conditions of

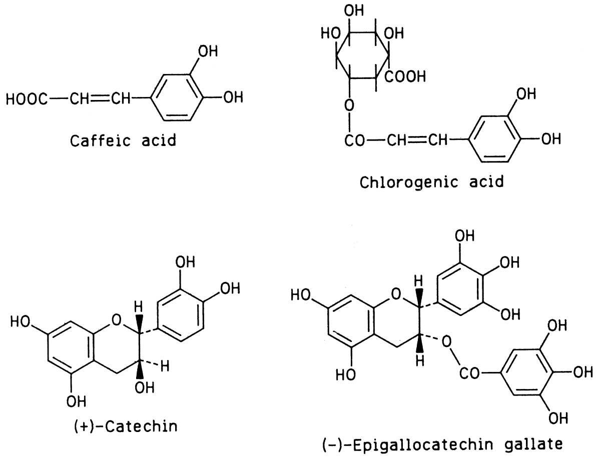


FIGURE 13 Four phenolic compounds involved in enzymatic browning.

the reactions. Among sugars, pentoses are more reactive than hexoses, and hexoses are more reactive than reducing disaccharides. When free amino acids react with sugars, lysine appears to be the most active among them. In peptides and proteins, the N-terminal amino acid is the most reactive, followed by a nonterminal lysine. Raising the temperature and/or the pH accelerates the Maillard reaction. Intermediate water activity appears to maximize this reaction.

Several pathways have been proposed for the formation of melanoidins through the interaction of carbonyl and amine compounds. A simplified scheme is shown in Fig. 14. This scheme involves first the condensation of a carbonyl compound (an aldohexose in this scheme) with an amine to a Schiff base via an intermediate product (not shown). The Schiff base quickly cyclizes to an N-substituted glycosylamine. Up to this step the process is reversible because the glycosylamine can be hydrolyzed back to the initial reactants. The N-substituted glycosylamine is then rearranged to an N-substituted 1-amino-1-deoxy-2-ketose, the Amadori compound, which is in equilibrium with its enol form. If the initial carbonyl compound is a 2-ketose (e.g., fructose), the corresponding N-substituted 2-amino-2-deoxy-1-aldose is formed by the Heyns rearrangement, which is analogous to the Amadori

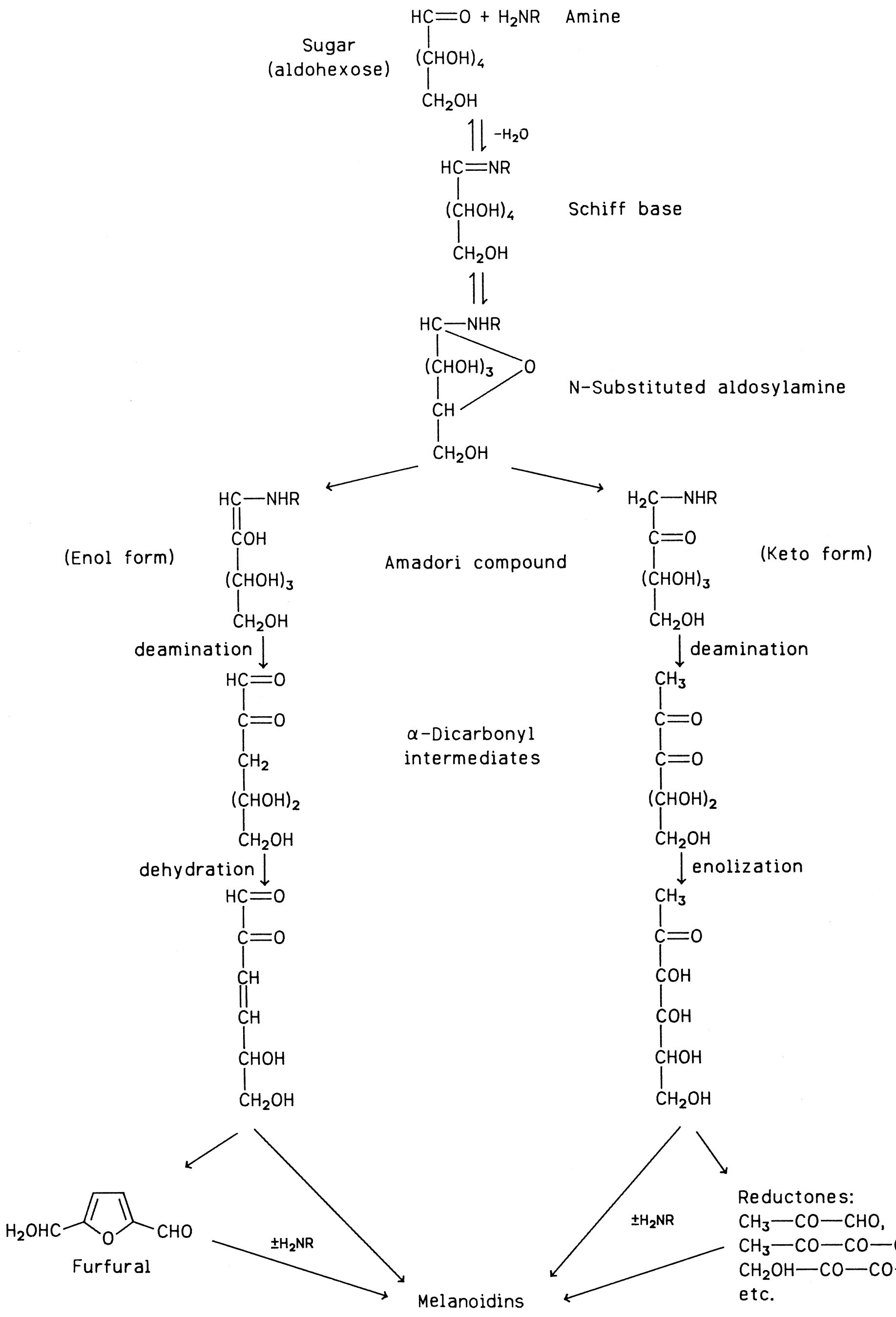


FIGURE 14 Simplified scheme of melanoidin formation by the sugar–amine reaction (± means presence or absence).

rearrangement for aldoses. The Amadori and Heyns compounds are subsequently subjected to a variety of transformations, which may include deamination, dehydration, enolization, cyclization, and degradation. The products of those reactions finally condense or polymerize with or without the participation of an additional amino compound to form dark pigments of a colloidal nature and ill-defined composition—the melanoidins. Some of the intermediate products are collectively called reductones, because they are strongly reducing compounds that account for the reducing properties of systems undergoing Maillard browning.

It should be mentioned here that a side reaction of the Maillard browning results in the formation of flavorful compounds, such as those associated with the roasting of meat, coffee, or nuts and the baking of bread. This

side reaction, known as the Strecker degradation, occurs between α-amino acids and dicarbonyl compounds and leads to the formation of aldehydes possessing one less carbon atom than the corresponding initial amino acids. The newly formed aldehydes are responsible for the pleasing flavors.

A significant consequence of the Maillard reaction is the loss of the nutritional value of the amino acid involved in the reaction. If the participating amino acid is essential, and especially if it is a limiting one, as is lysine in most cereal grains, the Maillard reaction can seriously lower the nutritive value of the food. Toasting, for example, may reduce to one-half the protein efficiency ratio of bread.

2. Caramelization

This reaction leads to brown products when sugars are heated dry or in solution. Certain conditions of caramelization favor the formation of flavor compounds as well. The chemical transformations involved in caramelization are complex and poorly understood. They include dehydration, fragmentation, and polymerization. On the heating of pentoses, furfural is formed which polymerizes to brown products. Heating hexoses results in hydroxymethylfurfural, which polymerizes similarly. The large quantities of industrial caramel color that are added to beverages (cola drinks), baked goods, and confections are made by heating high-conversion corn syrups in the presence of catalysts (acids, alkalis, salts).

3. Ascorbic Acid Browning

When ascorbic acid is heated in the presence of acids, furfural is formed. The latter, either by itself or after reacting with amino compounds, polymerizes to brown products. Citrus juices, especially their concentrates, develop browning, which has been attributed to ascorbic acid degradation.

4. Metal–Polyphenol Browning

Polyphenolic compounds form complexes with certain metals. The polyphenols of fruits and vegetables most commonly chelate iron. The resulting iron complexes are bluish black pigments. Cutting apples with a non-stainless-steel knife results in darkening of both the blade and the cut surface of the apple. This darkening is independent of the enzymatic browning that might develop as a result of cutting. Wine makers avoid contact between the wine and iron implements because of the black iron–tannin precipitate that forms on such contact. Certain varieties of potatoes tend more than others to darken after cooking. This darkening is attributed to a complex between iron and chlorogenic acid. The iron of the tissue must first be oxidized to the ferric state for the blackish complex to appear. The stem end, which contains much less citric acid than the rest of the tuber, displays the deepest darkening. Canned or pickled cauliflower may turn dark due to the interaction of polyphenols in the tissue with iron from external sources.

As already indicated, nonenzymatic browning is desirable in certain instances and undesirable in others. The availability of reactants and the type of conditions (temperature, pH, moisture) will determine the extent of browning. A chemical preservative often used to inhibit nonenzymatic (and enzymatic) browning is sulfur dioxide. An obvious way to prevent metal-polyphenol browning is to eliminate contact between susceptible tissues and reactive metals and use inoffensive equipment (stainless steel, glass-lined tanks, etc.)

IV. COLOR ADDITIVES IN FOODS

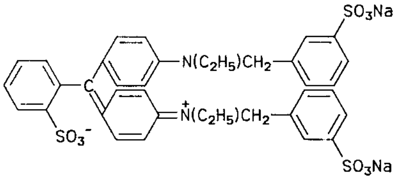
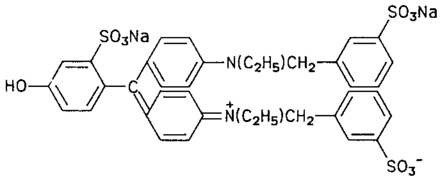
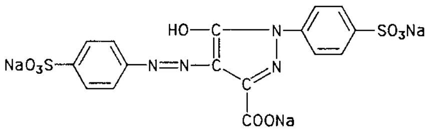
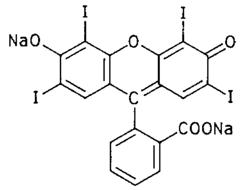
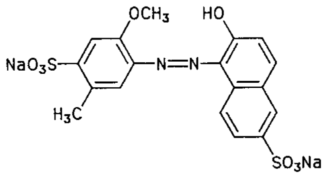
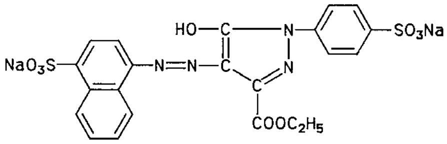
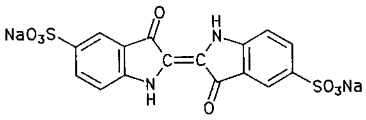
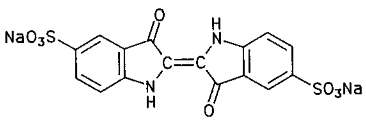
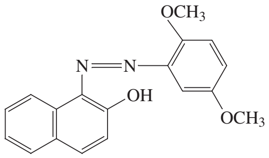
Colored substances are added to foods to modify the appearnce of insufficiently colored or discolored foods and to create new foods. The following criteria must be met if a food color (colorant) is to be used; (1) it must be safe at the level and under the conditions of use; (2) it must be stable in the products in which is added; (3) it must not impart any offensive property (flavor, texture) to the product; (4) it must be easy to apply; (5) it must have a high tinctorial power; and (6) it must not be too costly.

There are two classes of color additives, those that must be certified and those that are exempt from certification. Both are strictly controlled in the United States by regulatory statutes (Food Color Additives Amendments), but an official certificate is required for each commercial batch of color of the first group, while no such certificate is necessary for the second group. For certification the manufacturer must submit a sample of the batch to the Food & Drug Administration for chemical analysis. The results of the analysis are compared with the specifications for certified colors published in the Code of Federal Regulations. If the compliance is complete, a certificate is issued for that particular batch of color.

When a new color is to be introduced, the petitioner is expected to provide data proving that the new additive is safe and effective. The safety tests are elaborate and expensive. They include chronic toxicity feeding tests with two animal species over several generations. The effectiveness tests include long-term color stability experiments in the foods to which the color is to be added.

The food color additives subject to certification are listed in Table IV. The initials FD&C stand for the Food, Drug & Cosmetic Act under which these additives are regulated. The food color additives exempt from certification are listed in Table V. Generally, only synthetic organic

TABLE IV Permanently Listed Food Colors Subject to Certification

Color	Structure	Uses[a]
FD&C Blue #1		Used in amounts consistent with GMP
FD&C Green #3		Used in amounts consistent with GMP
FD&C Yellow #5		Used in amounts consistent with GMP
FD&C Red #3[b]		Used in amounts consistent with GMP
FD&C Red #40		Used in amounts consistent with GMP
Orange B		Coloring sausage surfaces or casings (150 ppm max. based on finished product)
FD&C Blue #2		Used in amounts consistent with GMP
FD&C Yellow #6		Used in amounts consistent with GMP
Citrus Red #2		In skins of oranges that are not intended for processing (at 2 ppm max., based on whole fruit)

[a] GMP. Good manufacturing practices.

[b] The FDA has recently banned the use of Red #3 in such products as cake frostings, certain processed foods, cough drops, and lipstick, in which the color is mixed with other additives reacting with it. This dye can still be applied directly to meat, nut products, fruit and fruit juices, candy, confections, and breakfast cereals.

TABLE V Permanently Listed Colors Exempt from Certification

Colorant	Uses[a]	Colorant	Uses[a]
Caramel	In foods, generally consistent with GMP	Cochineal extract; carmine	In foods, generally consistent with GMP
β-Carotene	In foods, generally consistent with GMP	Dehydrated beets	In foods, generally consistent with GMP
Annatto extract	In foods, generally consistent with GMP	Riboflavin	In foods, generally consistent with GMP
Paprika	In foods, generally consistent with GMP	Carrot oil	In foods, generally consistent with GMP
Paprika oleoresin	In foods, generally consistent with GMP	β-Apo-8′-carotenal	In foods, generally not to exceed 25 mg/lb
Turmeric	In foods, generally consistent with GMP	Titanium dioxide	In foods, generally not to exceed 1% by weight
Turmeric oleoresin	In foods, generally consistent with GMP	Grape skin extract	In still and carbonated beverages and alcoholic beverages
Saffron	In foods, generally consistent with GMP	Ferrous gluconate	For coloring ripe olives, consistent with GMP
Fruit juice	In foods, generally consistent with GMP	Canthaxanthin	In foods, generally not to exceed 30 mg/lb
Vegetable juice	In foods, generally consistent with GMP		
Toasted, partially defatted, cooked cottonseed flour	In foods, generally consistent with GMP		

[a] GMP, Good manufacturing practices.

colorants are subject to certification, while natural organic and inorganic colors, such as paprika and titanium oxide are not. The colorant β-carotene is not subject to certification whether it is obtained from a natural source or it is synthetically produced.

While synthetic food dyes are generally water-soluble, food lakes are water-insoluble. Food lakes are prepared by precipitating dyes on alumina. These lakes are useful for coloring water-repelling foods, such as fats and oils, certain gums, as well as packaging materials, e.g., plastic films, lacquers and inks, from which soluble dyes would leach out. Listing of a food dye does not necessarily imply listing the corresponding lake.

Polymeric food dyes have been developed that cannot pass the gastrointestinal wall and are excreted virtually intact in the feces. Toxicity and efficacy tests must be completed before FDA approval is granted to these dyes.

In recent years, plant tissue culture techniques have been applied to the production of food colors. Also the pigments of two fungi: *Monascus anka* and *Monascus purpureus* are being considered for use in foods. These fungal pigments have been used as food colors and medicines in the Far East for hundreds of years.

The regulations regarding color additives can be found in the Code of Federal Regulations, Title 21, Parts 70–82. Changes in these regulations are published in the "Federal Register." Additional information on color additives can be obtained from:

Food & Drug Administration
Division of Color & Cosmetics
200 C Street. S. W. Washington, DC 20204

SOME DOMESTIC SUPPLIERS OF COLOR ADDITIVES

Beatrice Foods Co., 156 W. Grand Ave., Beloit, WI 53511

BIOCON Inc., 518 Codell Dr., Lexington, KY 40509

COLORCON Inc., Moyer Blvd., West Point, PA 19486

Crompton & Knowles Co., 1595 MacArthur Blvd., Mahwah, NJ 07430

Hilton-Davis Co., 2235 Langdon Farm Rd., Cincinnati, OH 45237

H. K. COLOR Group, 155 Helen St., South Plainfield, NJ 07080

Hoffmann-LaRoche Inc., 304 Kingsland St., Nutley, NJ, 07110

Meer Corp., 9500 Railroad Ave., North Bergen, NJ 07047

Pylam Products Co., 1001 Stewart Ave., Garden City, NY 11530

Sethness Products Co., 2367 W. Logan Blvd., Chicago, IL 60647

Sun Chemical Corp., 441 Tompkins Ave., Staten Island, NY 10305

Warner-Jenkinson Co., 2526 Baldwin St., St. Louis, MO 63106

Whittaker, Clark & Daniels Inc., 1000 Coolidge St., South Plainfield, NJ 07080

SEE ALSO THE FOLLOWING ARTICLES

Biopolymers • Natural Antioxidants in Foods • Polymers, Synthesis

BIBLIOGRAPHY

Association of Official Analytical Chemists (1995). “Official Methods of Analysis” 16 ed., Washington, DC.

Food & Drug Administration (1982). “Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives in Food,” (The Redbook), Washington, DC.

Francis, F. J. (1986). “Handbook of Food Colorant Patents,” Food & Nutrition Press (1999). Westport, CT.

Hutchings, J. B. (1999). “Food Color and Appearance,” 2nd ed., Aspen Frederick, MD.

Ilker, R. (1987). “In-vitro pigment production,” *Food Technol*. **41**(4), 70.

Markakis, P., (ed.). (1982). “Anthocyanins as Food Colors,” Academic Press, New York.

Marmion, D. M. (1991). “Handbook of U.S. Colorants for Foods, Drugs and Cosmetics,” 3rd ed., Wiley-Interscience, New York.

Walford, J., (ed.). (1984). “Developments in Food Colours,” Vols 1 and 2, Elsevier, London.

Glycoconjugates and Carbohydrates

Eugene A. Davidson
Georgetown University School of Medicine

I. Monosaccharides
II. Oligosaccharides
III. Polysaccharides
IV. Glycoconjugates
V. Analytical Methods

GLOSSARY

Aglycone Group attached to the hydroxyl of a furanose or pyranose sugar to form a glycoside. Can vary from a simple methyl group to other sugars to complex alkaloids.

Aldose Polyhydroxy aldehyde with three or more carbon atoms. When of sufficient length will form five- or six-membered ring (furanose or pyranose) hemiacetals. Designated D- or L- based on the configuration of the asymmetric center furthest from the carbonyl group relative to glyceraldehyde.

Anomer Formation of the cyclic ring structure confers asymmetry on the original carbonyl carbon atom. The two anomers produced are alpha- (the newly formed hydroxyl group projects on the same side as the five- or six-membered ring) or beta- (the new hydroxyl projects on the opposite side of the newly formed ring). This is combined with D-, or L-; α-D-glucopyranose, β-L-fructofuranose, for example.

Epimer Saccharide that differs from a reference sugar at a single asymmetric center. Thus, D-galactose is the 4-epimer of D-glucose.

Furanose Five-membered ring (derived from furan) with a single oxygen formed by reaction of the carbonyl carbon of an aldose or ketose with the appropriate hydroxyl group. Sugars commonly found as furanoses include ribose and 2-deoxyribose (in RNA and DNA, respectively), and fructose in combined form (sucrose, for example).

Glycolipid Structure containing one or more saccharide units covalently attached to the primary hydroxyl group of *N*-acyl sphingosine (ceramide), a C-18 amino alcohol. Those containing sialic acid are termed *gangliosides*.

Glycoprotein Protein with covalently attached saccharides. Linkage may be *N*- (amide nitrogen of asparagine residues), or *O*- (serine and threonine are most common). The saccharide units may be heterogeneous within the same protein, even at the identical substitution site.

Glycosaminoglycan Polysaccharide chain found in

proteoglycans. Has a linear structure that features a repeating disaccharide unit comprised of glucosamine or galactosamine and a uronic acid or galactose. The amino sugars are generally *N*-acetylated and contain ester sulfate. The pattern of sulfation varies and may confer specific biological properties.

Glycoside Acetal or ketal formed by substitution of the anomeric hydroxyl with an aglycone. This reaction fixes the configuration at the anomeric center. Linkages are generally acid labile.

Ketose (ulose) Polyhydroxy ketone (at carbon 2) of four or more carbon atoms (see *Aldose*).

Mutarotation The change in optical rotation associated with the equilibration, in solution, of the anomeric and ring forms of a saccharide beginning with a single crystalline form. Thus, dissolving α-D-glucopyranose in water allows formation of the β-anomer as well as the furanose forms. Since all have different optical rotations, the initial value changes until the equilibrium mixture is achieved.

Oligosaccharide Sugars that contain 2–10 monosaccharide units. These may be linear or branched. The most common representatives are derived from aldoses and will contain a single reducing terminus (free, or potential, aldehydo group) and one or more nonreducing termini (linked to other sugars only at their anomeric hydroxyl group).

Polysaccharide Structure containing more than 10 monosaccharide units. These may be the same (homopolysaccharide) or different (heteropolysaccharide), and the structure may be linear or branched. Natural diversity is very large.

Proteoglycan Special category of glycoproteins in which the saccharide chain is covalently linked to protein via an initiating xylose residue, a specific core oligosaccharide, and is followed by a glycosaminoglycan chain.

Pyranose Six-membered ring form of a saccharide formed by reaction of the appropriate hydroxyl group with the carbonyl carbon. Generally the most stable structure in solution. The conformer resembles cyclohexane and is generally a chair form with axial and equatorial substituents.

THE UTILIZATION of carbohydrates as food sources, sweetening agents, and clothing materials dates back several thousand years. Likewise, the production of beer and wine was known in ancient times although the relationship to saccharides and fermentation was not appreciated then. Manufacture of sugar as a sweetener including refinements such as decolorization with charcoal also has ancient origins. The commercial importance of carbohydrates extends beyond the food industry to textiles, pharmaceuticals, and a wide range of chemicals.

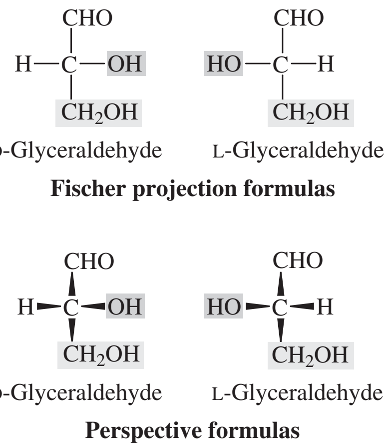


FIGURE 1 Projection formulas of D- and L- glyceraldehyde.

The common definition of carbohydrates is polyhydroxy aldehydes or ketones (the carbonyl group is generally at C-2), their derivatives, oligomers and polymers. The term itself arises from the empirical formula of the compounds initially studied (glucose, for example) since that is represented by $C(H_2O)_n$.

An appreciation of the diversity available in such structures began about 1900 with the work of van't Hoff on the tetrahedral carbon atom, extrapolated by Emil Fischer in his elucidation of the stereochemistry of glucose. During that series of experiments, Fischer was able to deduce the relative positions of all of the hydroxyls at the asymmetric centers and could relate glucose to glyceraldehyde (Fig. 1). However, he was unable to provide an absolute stereo structure for glyceraldehyde and thus had to decide between mirror image structures for natural glucose. That he chose the correct one may reflect little more than that he was right handed. He did synthesize all of the possible aldohexoses.

I. MONOSACCHARIDES

It is generally accepted that monosaccharides, or simple sugars, may have three to nine carbon atoms in a linear chain. Thus the nomenclature aldotriose, aldotetrose, aldopentose, etc., or ketotetrose, ketopentose, etc. The vast majority of naturally occurring sugars contain five or six carbon atoms. These saccharides are classified as D- or L- if they can be derived from D- or L-glyceraldehyde, respectively, or the four carbon ketoses (Fig. 2). The determining asymmetric center is that farthest from the carbonyl group—projection of the hydroxyl group to the right is D. As the chain length increases, new asymmetric centers

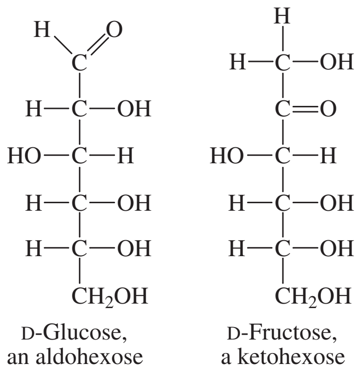


FIGURE 2 Projection formulas of D-glucose and D-fructose.

are introduced. Accordingly, there are four aldotetroses, eight aldopentoses, four ketopentoses, eight ketohexoses, etc. (Fig. 3).

Fischer recognized that sugars such as glucose existed in a ring form (i.e., a hemiacetal or hemiketal formed by addition of a hydroxyl group to the carbonyl carbon). Based on his knowledge of lactones, he assigned the participating hydroxyl to C-4 in the case of glucose, thus forming a five-membered (furanose) ring. The thermodynamics, however, show that the six-membered (pyranose) ring is the more stable form and that is the one predominating in solution. Note that the internal addition reaction at C-1 causes that carbon to become asymmetric and, therefore, two new isomeric structures are formed. To distinguish this asymmetry from that present at the other asymmetric centers, the term *anomer* is employed with the designation α or β reflecting projection of the newly formed hydroxyl group on the same or opposite side of the ring, respectively (Fig. 4).

The ring structures (aldopyranose for the six-membered rings and aldofuranose for the five-membered rings) are, in aqueous solution, in ready equilibrium with the open-chain, free aldehyde form. Thus, a solution of glucose in water contains five species: free aldehyde, two furanoses, and two pyranoses (Fig. 5). The aldehyde form represents about 0.025% of the total, with the bulk made up of the two pyranoses (the beta form predominates; see below). Although the six-membered ring is the energetically preferred form, the predominant ketohexose, D-fructose, is found in the furanose form in combined structures (sucrose, for example) although it too prefers the six-membered ring when free in aqueous solution. Other commonly occurring furanosides include ribose and deoxyribose as components of RNA and DNA (Fig. 6); galactose is also found as a furanoside in several plant and microbial polysaccharides.

A key property of the free monosaccharides is their ability to be readily oxidized (initially via the available aldehydo function), especially in alkaline solution. This characteristic was used as an analytical tool for many years, and the term *reducing sugar* was employed to designate saccharides with a free (or potentially free, in aqueous solution) carbonyl function. Prior to the advent of specific enzymatic methods, this was the standard procedure for measurement of blood glucose levels (i.e., reducing sugar).

Monosaccharides that differ from one another at a single asymmetric center other than the one derived from the carbonyl carbon are termed *epimers* (D-galactose is the 4-epimer of D-glucose, and D-mannose is the 2-epimer, for example; Fig. 7). Mirror-image structures are termed *enantiomers*.

An important and characteristic feature of saccharides is their ability to interact with plane-polarized light. Thus, each sugar has a characteristic optical rotation that is a complex function of the asymmetric centers, their polarizability, solvent interactions, etc. The sign (+ or dextrorotatory, − or levorotatory) and magnitude cannot be determined *a priori*. Nor is it true that all sugars of the same configuration (D- or L-) will have the same sign of optical rotation. Thus, D-glucose is dextrorotatory, whereas D-fructose is levorotatory. Since monosaccharides tend to crystallize in a single anomeric form rather than as mixtures, dissolving them in water results in a change in the optical rotation from that of the pure anomer (either alpha or beta) to that of the thermodynamically defined equilibrium mixture. This phenomenon is termed *mutarotation* (Fig. 8).

As the principles of conformational analysis became known, it was rapidly realized that the pyranose form of sugars adopted a chair structure analogous to that of cyclohexane (Fig. 9). In this spatial arrangement, equatorial hydroxyl groups are thermodynamically more stable (energy differences of about 1.5 kcal) than axial ones with the exocyclic hydroxymethyl group having a much larger effect. It is not surprising, therefore, that β-D-glucose, the all-equatorial structure, is the most prevalent natural sugar and the most prevalent organic compound on earth (Fig. 10). As a philosophical aside, it may be inferred that the choice between D- and L-glucose was made very early and on a basis we do not understand since they are conformationally equivalent. It is also expected that other naturally occurring sugars such as D-mannose or D-galactose would have only a single axial hydroxyl. Of the 16 possible aldohexoses, D-glucose, D-mannose, and D-galactose are widely distributed in nature while the remainder are of laboratory interest only. Parenthetically, idose (three axial hydroxyls in the classic conformer) has never been obtained in crystalline form. Additional widely distributed sugars include D-xylose (all-equatorial aldopentose), D-ribose, and 2-deoxy-D-ribose (backbone components of RNA and DNA,

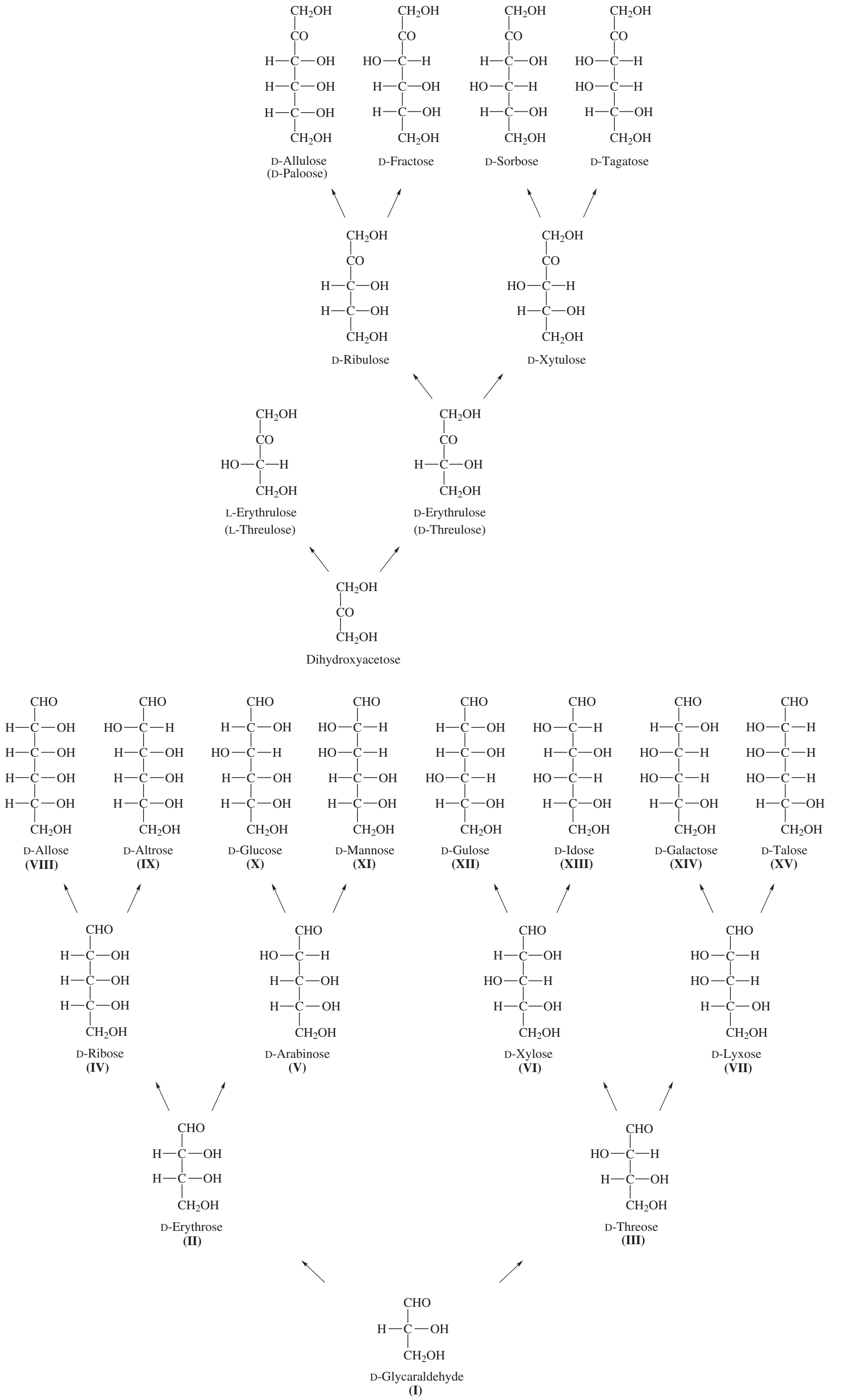


FIGURE 3 Structures of aldoses and ketoses up to six carbons in length. Monosaccharides with up to nine carbons are present in nature.

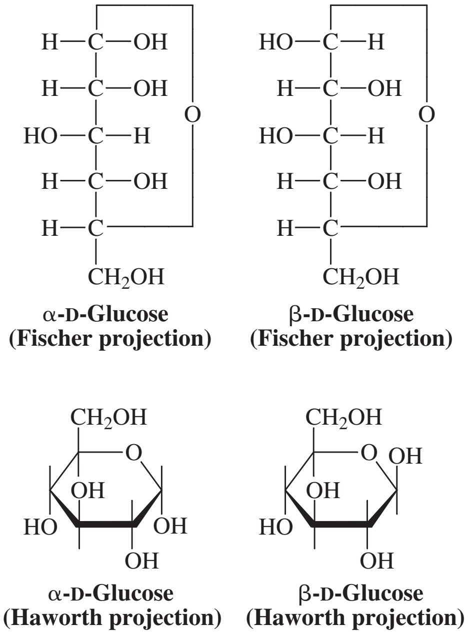


FIGURE 4 Alpha- and beta-forms of D-glucose.

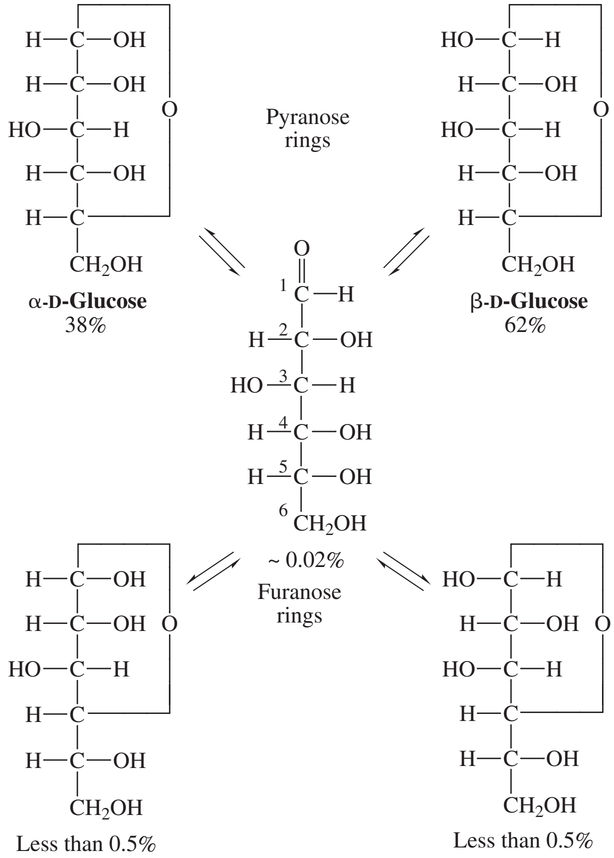


FIGURE 5 Equilibrium mixture of D-glucose in aqueous solution.

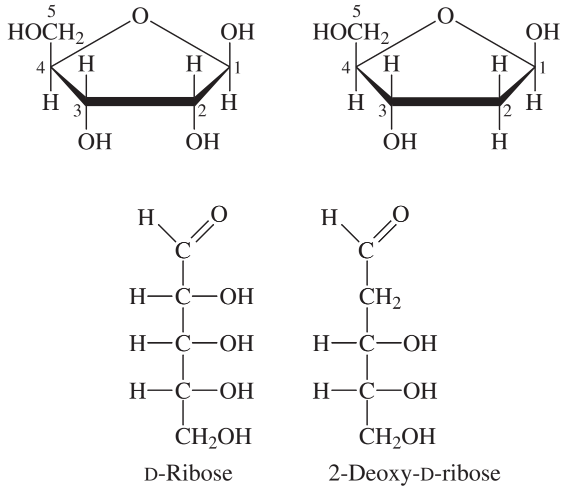


FIGURE 6 Structures of D-ribose (left) and 2-deoxy-D-ribose (right).

respectively). A large number of other sugars are present in nature. The vast majority of naturally occurring saccharides exist other than as free monosaccharides (only D-glucose and D-fructose are widely found free) and are present in disaccharides and complex oligomeric and polymeric structures in both plants and animals. About a dozen different sugars are present in mammals, all but glucose in combined form. Plants and microorganisms have very diverse saccharides, again all in combined forms.

A. Derivatives—Natural and Laboratory

The hydroxyl group formed as a result of ring closure represents a site for attachment of a broad variety of substituents. Compounds that are formed in such reactions are full acetals (ketals) and thus no longer undergo interconversion at the anomeric center. The configuration of such glycosides is therefore either alpha or beta depending on the relationship between the C-1 group and

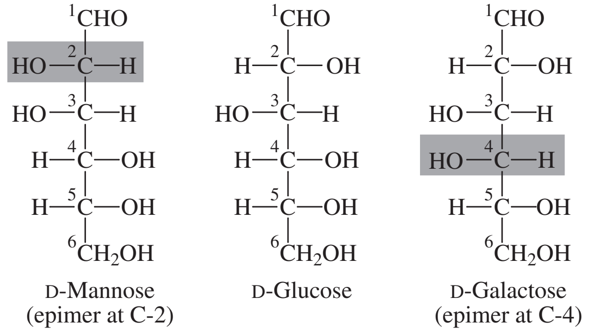


FIGURE 7 Relationship between D-glucose and D-galactose (4-epimer), and D-mannose (2-epimer).

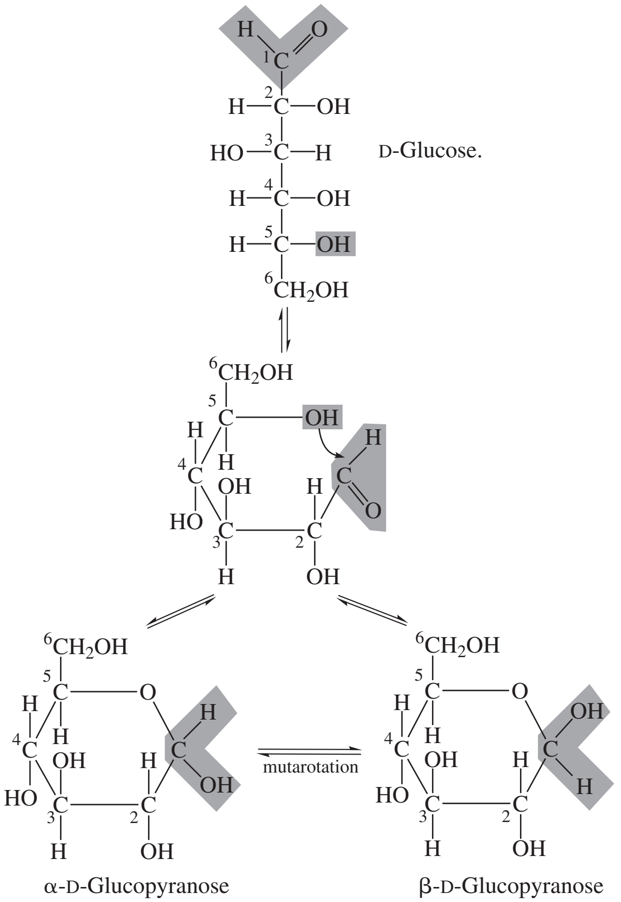


FIGURE 8 Interconversion of glucose forms in aqueous solution. The optical rotation is initially reflective of the starting material (alpha or beta) and changes (mutarotates) until equlibrium is achieved.

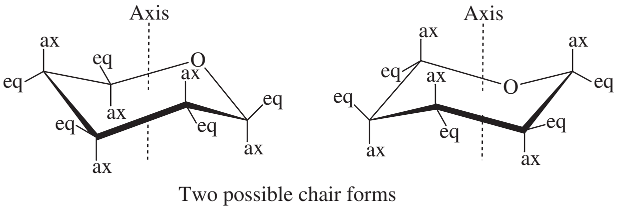


FIGURE 9 Conformational structure of pyranose rings.

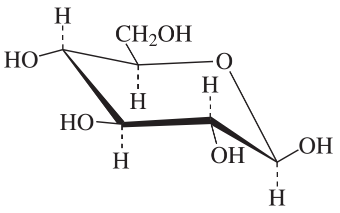


FIGURE 10 All-equatorial structure of β-D-glucose.

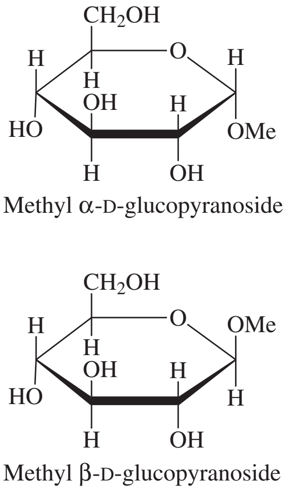


FIGURE 11 Methyl glucopyranosides.

the projection of the ring; if on the same side, designate alpha, otherwise, beta (Fig. 11). A wide variety of natural and man-made derivatives (glycosides) are known with the substituents, aglycones, varying from simple methyl groups to complex organic molecules including other sugars (see below).

In addition to substitution of the anomeric hydroxyl, many modifications of the hydroxyl loci are known in nature. Most prevalent are those in which the hydroxyl group at C-2 is replaced by an amino function, generally acetylated. The sugar 2-deoxy 2-acetamido-D-glucose (*N*-acetylglucosamine) is distributed throughout nature and, in its polymeric form (chitin), forms the organic matrix of insect and arthropod exoskeletons. Hence, it is likely the second most prevalent organic molecule on earth. Other variations include oxidation (C-6 or C-1) to form carboxyl groups and loss of a hydroxyl to form deoxy sugars (Fig. 12).

B. Chemical Transformations

The chemistry of carbohydrates involves transformations at both the carbonyl function and the chain hydroxyl groups. The presence of multiple species of simple sugars in aqueous solution gives rise to complex reaction patterns wherein product distribution may well be determined by kinetic rather than thermodynamic considerations.

1. Reactions at the Carbonyl Group

The carbonyl function reacts directly with a variety of reagents either as the free aldehyde or keto group or as the corresponding hemiacetal or hemiketal. In early work, Fischer introduced phenylhydrazine, which condenses at

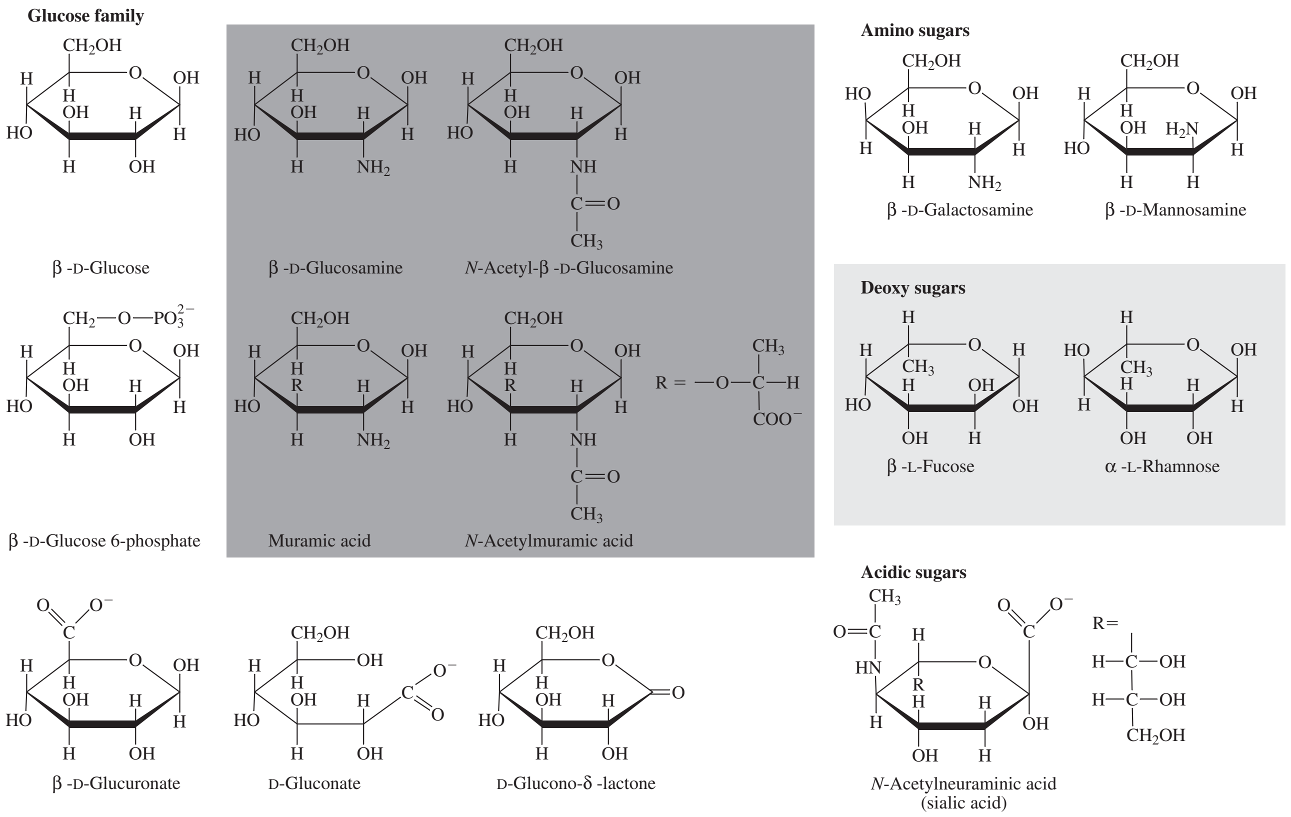


FIGURE 12 Sugars related to D-glucose that occur in nature.

the carbonyl carbon and oxidizes the adjacent secondary hydroxyl (aldoses) or primary alcohol (ketoses) to allow a second condensation. The resulting derivatives, phenylosazones, are crystalline and were employed both for identification and determination of stereochemistry. Phenylhydrazine itself is a liver poison and it is generally believed that Fischer suffered from liver damage due to this reagent (no longer employed). A simple example of the utility of this conversion is the fact that D-glucose, D-mannose, and D-fructose all give the same phenylosazone showing that their relative stereochemistry at the remaining asymmetric centers (carbons 3,4,5) is identical (Fig. 13).

Certain transformations may dictate that the carbonyl function remain available for later chemistry while reactions are carried out elsewhere in the chain. This may be achieved by formation of dithioacetals or, in the ring form, by suitable substitution of the anomeric hydroxyl group (see below).

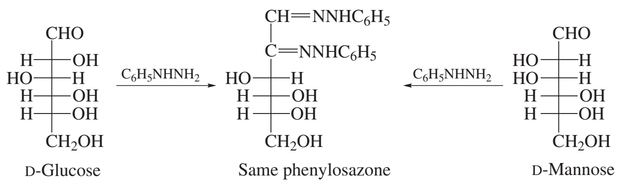


FIGURE 13 Formation of phenylosazone from D-glucose or D-mannose. The loss of asymmetry at C-2 yields identical products.

2. Reaction at the Anomeric Hydroxyl Group

The hemiacetal nature of the anomeric hydroxyl group makes it the most reactive of that type. Direct oxidation can be carried out with several reagents, most classically, hypoiodite. This results in rapid oxidation to the aldonolactone and is only exhibited by aldoses. Alternatively, reaction with alcohols results in the formation of full acetals (glycosides). A very large variety of such structures have been made.

Control of stereochemistry at the anomeric center is complex, involves the configuration at carbon two and kinetic factors. In general, both alpha- and beta-glycosides are formed. The configuration is very important in biological systems since enzymatic transformations are generally stereospecific with essentially all glycosidases having absolute specificity for either the alpha or beta form. Glycoside formation is generally carried out by reaction of the sugar with the appropriate alcohol in nonaqueous solution, typically with an acid catalyst.

3. Reactions at Secondary Hydroxyls

Substitution reactions at secondary hydroxyls are generally performed either for analysis of structure or to serve a protective function during other reactions. Etherification of the nonanomeric hydroxyls was an important structural tool in the analysis of oligosaccharide and polysaccharide structure. Methyl ethers have been employed for structural determination for more than 75 years. Thus, methyl ether formation in a polysaccharide results in substitution only at free hydroxyls. Subsequent analysis of the methylated derivatives reveals positions previously occupied in glycosidic linkage. Reagents used for this purpose have evolved from dimethylsulfate to the commonly employed method of Hakomori using sodium hydride and dimethylsulfoxide.

Another frequently used ether substituent is the benzyl group, which is stable under a variety of reaction conditions but can be removed by catalytic hydrogenation. This is employed mainly during synthetic schemes where protection of specific hydroxyls is required. Currently, ether formation is used primarily as an adjunct to mass spectrometric analysis.

The secondary hydroxyl groups can also be esterified. Acetyl substitution via acetic anhydride is often used, especially as a protective group. Tosyl substitution using toluenesulfonyl chloride is often employed since the properties of that ester (good leaving group) allow for SN_2 displacement resulting in inversion of configuration. Acetate esters are readily removed under basic conditions (methoxide ion) when glycosides remain intact.

4. Reactions at the Hydroxymethyl Group

The exocyclic nature of this group makes it the second most reactive of the hydroxyl functions. Tritylation is often employed to transiently block C-6 (ready removal under mild acid conditions). This is also the site of enzymatically catalyzed phosphorylation of glucose and other sugars, the first step in their metabolic utilization. Oxidation to a carboxylic acid is common. The resulting uronic acid is, for the common sugars, widely distributed in natural polysaccharides in both plants and animals. It is of interest that L-iduronic acid (the 5-epimer of D-glucuronic acid) and L-guturonic acid (the 5-epimer of D-mannuronic acid) are formed in nature after the respective precursor uronic acid has been incorporated into polymeric linkage.

5. Hydroxyl Pairs

The proximity and defined stereochemistry of hydroxyl groups in the ring structures of saccharides allows for a number of selective reactions. These include formation of acetals or ketals and of a group of interactions restricted to hydroxyls on adjacent carbons.

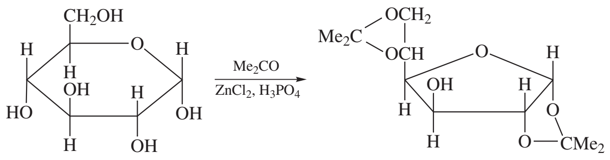


FIGURE 14 1,2 5,6-diisopropylidene D-glucofuranose. Formed by reaction of D-glucose with acetone in the presence of a suitable catalyst. The furanose product dominates due to kinetic control of the reaction.

Reaction of glucose with acetone results in the formation of 1,2 5,6-diisopropylidene glucofuranose (Fig. 14). The furanose product arises because of the high reactivity of the exocyclic hydroxymethyl group thus favoring the formation of the five-membered ring. This is a commonly used protective scheme since the ketal function is readily cleaved by mild acid; in fact, selective cleavage of the 5,6 ketal can be accomplished. Acetals are likewise formed; reaction of glucose with benzaldehyde results in formation of the 4,6 benzylidene derivative, another useful synthetic intermediate.

Adjacent hydroxyls that are *cis* can undergo several different reactions. Included in this group is complexation with borate to form a transient cyclic adduct (Fig. 15). This alters chemical reactivity and electrophoretic properties of the reactive sugar and has several applications.

A key property of *cis*-hydroxyls is their susceptibility to carbon–carbon bond cleavage by metaperiodate. This oxidative reaction proceeds via a five-membered ring intermediate, generates a pair of aldehydes, and results in scission of the carbon chain (Fig. 16). It has been demonstrated that the *cis* orientation is necessary for this reaction since fixed ring structures with *trans*-hydroxyls do not react. The reaction has been utilized to identify saccharides in tissues via subsequent treatment of the generated aldehydes with a suitable amine, forming a colored Schiff base product (periodate-Schiff reaction).

C. Biosynthesis

The fixation of carbon dioxide via photosynthesis is the initiating reaction in saccharide synthesis in nature. Light

FIGURE 15 Borate ester formation. *cis*-Hydroxyls are preferred. The extent of reaction was originally monitored by following the change in conductivity of borate solutions on the addition of saccharide.

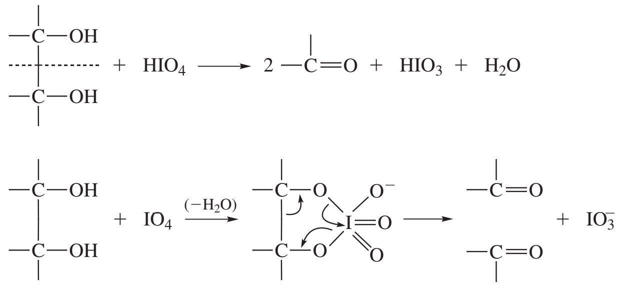


FIGURE 16 Action of periodate on vicinal diols. The proposed cyclic intermediate suggests that rigid structures with *trans*-hydroxyls will react poorly, a prediction confirmed experimentally.

energy is harvested via chloroplasts and used to provide chemical potential in the form of adenosine triphosphate, and reducing equivalents. The key intermediate, D-ribulose 1,5-bisphosphate, fixes carbon dioxide (the dark reaction) yielding products that are ultimately converted to D-glucose via a series of reactions of phosphorylated sugar intermediates. All other naturally occurring sugars are derived from glucose in transformations that involve phosphorylated or nucleotide-linked sugars. Thus, glucose-6-phosphate is converted to fructose-6-phosphate, which in turn is converted to mannose-6-phosphate; fructose-6-phosphate is also aminated to form 2-deoxy 2-amino glucose-6-phosphate. Galactose is formed by epimerization at C-4 of uridine diphosphoglucose (Fig. 17), fucose by a series of reactions initiating with guanosine diphosphomannose, etc. Thus, the diversity in saccharides seen in the biosphere stems from a single precursor, D-glucose. This is, therefore, the only required dietary saccharide for man.

II. OLIGOSACCHARIDES

The ability of the anomeric hydroxyl group of a sugar to be substituted with another sugar (via one of its hydroxyl groups), and for this process to be iterated leads to formation of di- and higher saccharides. Because several hydroxyls are available for such linkages and the anomeric configuration can vary as well, the number of possible structures grows exponentially even for oligomers derived from the same sugar. This may explain, in part, why many biological recognition events involve saccharides as a ligand.

Two disaccharides are widely distributed in nature: sucrose and lactose. Sucrose (α-D-glucopyranosyl-β-D-fructofuranoside) is the table sugar of commerce and a major industrial product (Fig. 18); primary sources are sugar cane and beets. This saccharide and its source material (cane molasses) serve as the basis for rum production; cruder precursors are important medium additives for the

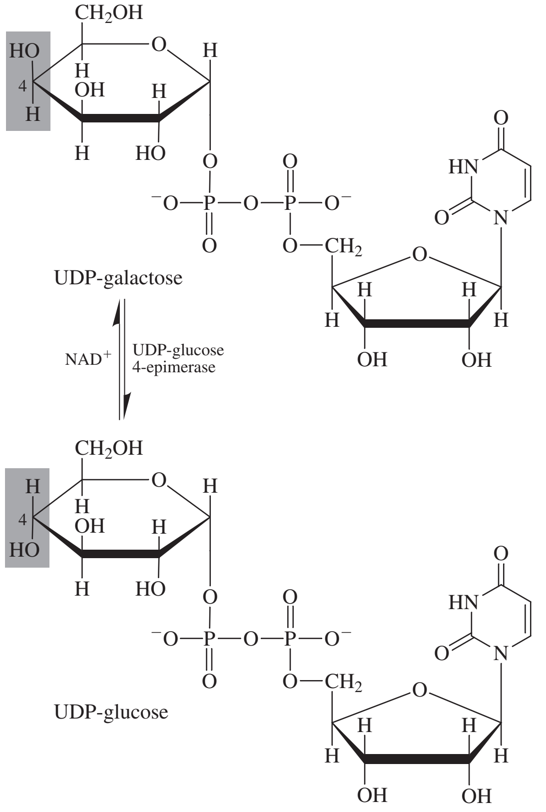


FIGURE 17 Biosynthesis of D-galactose by epimerization of uridine diphosphoglucose.

production of antibiotics. The sweetness of sucrose, and other simple sugars, is a major aspect of their commercial importance. The fact that sucrose is a nonreducing sugar contributes to its stability and market value.

Lactose (β-D-galactopyranosyl-1-4-D-glucopyranose) is the major sugar of milk (Fig. 19). It is of interest that the ability to utilize either sucrose or lactose as a source of calories is dependent on their enzymatic hydrolysis to the constituent monosaccharides in the intestine. About one-third of the oriental population lacks the requisite galactosidase and is thus lactose intolerant; this same problem occurs in some infants (developmentally related) and results in a "colicky" baby.

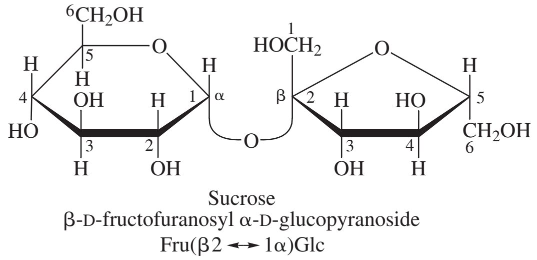
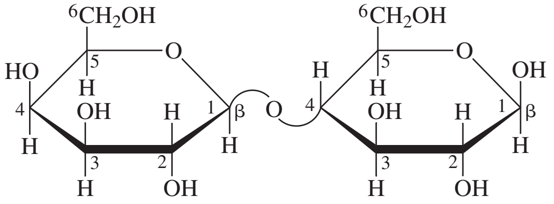


FIGURE 18 Structure of sucrose. Note that this sugar is nonreducing.



Lactose (β form)
β-D-galactopyranosyl-(1 → 4)-β-D-glucopyranose
Gal(β1 → 4)Glc

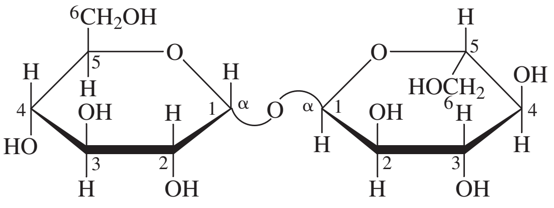
FIGURE 19 Structure of lactose, the major sugar present in milk.

Other naturally occurring disaccharides of interest are maltose (α-D-glucopyranosyl-1-4-D-glucopyranose), an intermediate in the digestion of starches, and trehalose (α-D-glucopyranosyl-α-D-glucopyranoside), the major sugar of insect blood (Fig. 20).

Naturally occurring tri- and higher oligosaccharides (less than 10 sugar units) are rare. Raffinose (3), stachyose (4), and verbascose (5) are all assembled on a sucrose core with substituents originating on the glucose moiety.

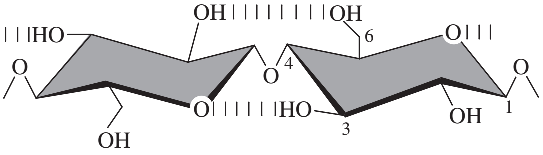
III. POLYSACCHARIDES

Polymeric saccharides can have 10,000 or more sugar units, four or five different sugars, and a branched structure. The diversity that can be envisioned is huge and much of it is actually observed. Classification of polysaccharides differentiates homopolymers (a single sugar type) and heteropolymers (two or more sugar types) with subtypes reflecting linear or branched structures within each group. The physical and biological properties of each polymer depend on the architecture of the molecule as well as the enzymatic machinery available to interact with it. Thus, cellulose, a linear β-1-4 glucose polymer is a stable, highly organized polymer, indigestible by man, whereas starch, its α-linked counterpart is a major food source for all mammalian species.



Trehalose
α-D-glucopyranosyl α-D-glucopyranoside
Glc(α1 ↔ 1α)Glc

FIGURE 20 Structure of trehalose, the predominant sugar present in the blood of insects.



(β1 → 4)-linked D-glucose units

FIGURE 21 The repeating unit of cellulose showing hydrogen bond interactions. The extended structure allows chains to stack via the relatively hydrophobic axial faces of the pyranose rings.

A. Homopolysaccharides

The major homopolysaccharides are cellulose, chitin, starches (amylose and amylopectin), glycogen, and xylans.

Cellulose, a linear glucose polymer linked β-1-4 is the predominant natural product in the biosphere. The all-equatorial structure allows for extensive hydrogen bonding, whereas the axial, somewhat hydrophobic faces favor a nonaqueous environment. The resulting aggregates have a highly ordered, quasi-crytalline arrangement (Fig. 21); hence, the unusual stability exhibited by the molecule, exemplified in trees and wooden artifacts. The broad distribution and low cost of cellulose has made it a major starting material for industrial development while the unmodified polymer remains the basis for wood, paper, and cotton.

Chitin, identical in linkage to cellulose but composed of N-acetylglucosamine instead of glucose, is the major structural component of insect and crustacean exoskeletons as well as a cell wall component of molds and fungi. The structural comments regarding cellulose also apply generally to chitin, especially in terms of stability. Less industrial development has been done with this polymer, in part because shrimp shells may present a more difficult starting material than trees.

Amylose and amylopectin are closely related, α-1-4-linked glucose polymers. They are major constituents of starches and hence key nutrients worldwide. Amylose is a linear chain with up to a few hundred glucose units. Amylopectin has the same backbone chain but α-1-6 branches approximately every 20 glucose units. The branches have the same linkages as the main chain. This ramified structure allows for efficient packing in cells.

Glycogen (Fig. 22), present in all higher animals, is closely related to amylopectin in that it has the same fundamental structure of a linear glucose chain with branches, and the same linkages. In this case, however, branches occur about every seventh residue, yielding a highly re-branched, tree-like envelope. This is essential for both packing in cells and for the rapid degradation of the polymer to provide critical metabolic intermediates. Glycogen serves as a primary energy reservoir in muscle and as a source of circulating glucose in the liver. It is of interest

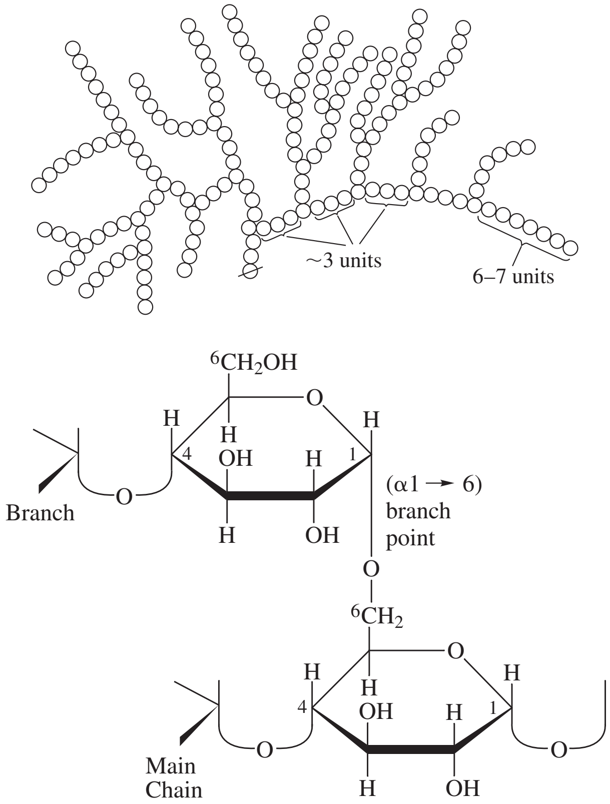


FIGURE 22 Schematic structure for glycogen and details of a typical branch point. The same linkages are present in amylopectin, which has less frequent branches.

that the biosynthesis of glycogen initiates on a core protein (glycogenin), which may be cleaved from the polysaccharide subsequent to polymerization.

Xylans are a group of polymers based on a structure analogous to that of cellulose wherein xylose is the repeating unit. The simplest representative contains only D-xylose with β-1-4 linkages and is a common component of plant walls. Several heteropolysaccharides utilize the xylan backbone and have various other saccharides as branches. Xylans are often associated with cellulose in plant cell walls.

B. Heteropolysaccharides

The diversity of heteropolysaccharides is enormous. Distributed throughout the animal and plant kingdom, this class of macromolecules rivals proteins in diversity. Initially thought to function only as structural components, it is now realized that specific sequences may have information content and can serve as recognition or signaling elements. In addition, modifications of saccharides already incorporated into a polymeric chain (sulfation, for example) add an additional level of complexity. The ability of these materials to function as other than strictly physical components necessitates, in many cases, that the saccharide chain have a defined three-dimensional conformation. Much work has been done to model the three-dimensional structure of proteins and many protein structures have been determined by X-ray crystallographic analysis. Much less has been done with complex saccharides in terms of either molecular dynamics or conformational analysis of larger structures. The following discussion is intended to be representative only.

In addition to cellulose and xylan, plant cell walls contain a variety of complex heteropolysaccharides including other glucans. Detailed structures have only been determined for a limited number of these structures and their interactions, possibly covalent, with other plant components have rarely been determined in detail. Commercially important are pectins, polygalcturonides, which are key components of jellies and related uronides (gums) used as thickening agents in a broad variety of applications ranging from ice cream manufacture to the production of printing ink. Agars are galactans that are sulfated and also contain 3,6-anhydro-L-galactose, which have gelation properties. They have broad application in microbiology.

Bacteria likewise produce a wide variety of complex, cell-surface polysaccharides. Strains of *Streptococcus pneumoniae*, the causative agent of bacterial pneumonia, are characterized by a capsular polysaccharide—more than 100 different types are known. Since the capsular material is immunogenic and protective, polyvalent vaccines have been developed that utilize the capsular polysaccharides from common strains. Similarly, organisms responsible for bacterial meningitis have a characteristic polysaccharide capsule that is also immunogenic and protective. Conversely, several bacteria have, as an exterior capsule, saccharides sufficiently similar to those produced in their animal host so as to serve as a mechanism for avoidance of host immune responses. Many gram-positive bacteria have, as essential cell wall components, a complex saccharide structure that contains muramic acid (a condensation product of *N*-acetylglucosamine and pyruvate) and other amino sugars and is cross-linked by a peptide (Fig. 23).

A widely distributed heteropolysaccharide, hyaluronic acid, has both commerical and biological importance. This molecule, a repeating structure of D-glucuronic acid and *N*-acetylglucosamine with β-1-3 and β-1-4 linkages, is found in bacterial and animal sources, and it is one of the few complex saccharides not covalently linked to protein (Fig. 24). Molecular weights vary depending on source but often exceed two million. The polyanionic nature of the molecule leads to a relatively extended solution conformation. This coupled with the highly hydrophilic chemistry results in solutions with very high viscosity, an important physical property. This is utilized in treatment of

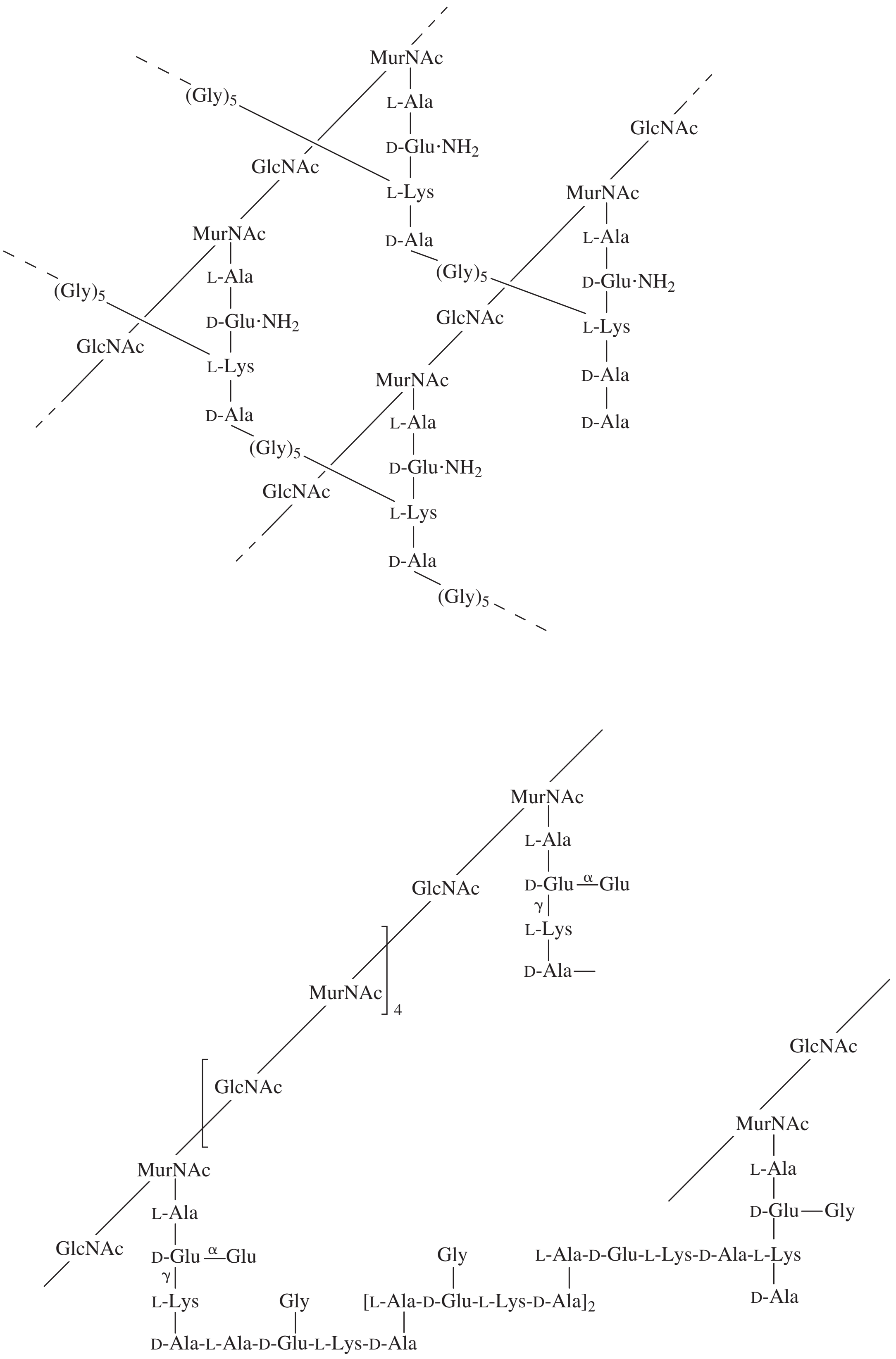


FIGURE 23 Typical bacterial cell wall structures (peptidoglycans).

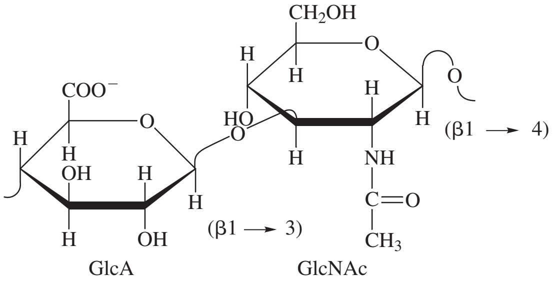


FIGURE 24 Repeating unit of hyaluronate. This polysaccharide is distributed throughout connective tissue and is the only mammalian polysaccharide not covalently attached to protein.

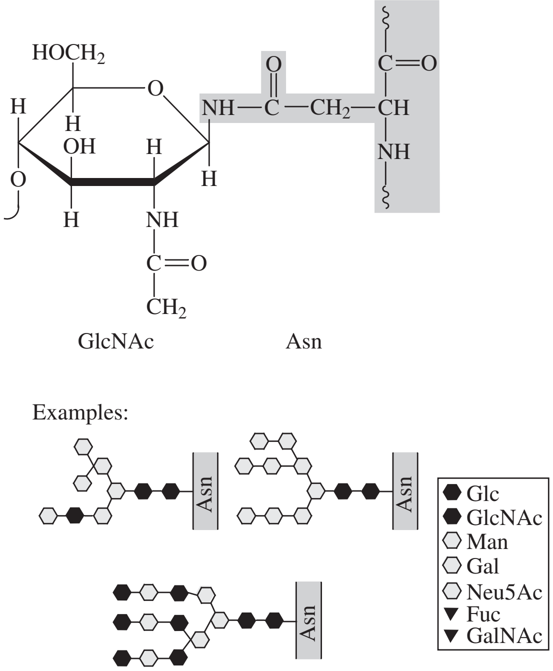


FIGURE 25 Schematic of a typical *N*-linked oligosaccharide. Note the core structure, which contains two *N*-acetylglucosamine and three mannosyl residues. This is present in all units of this type.

osteoathritis of the knee (an injectable) and in eye surgery as a viscoelastic. In addition, cell surface receptors have been identified that recognize the saccharides in hyaluronate, and interaction with specific proteins is responsible for the aggregate properties of connective tissue proteoglycans (see below). This diversity illustrates that a single polysaccharide may have both informational and physical roles in nature.

IV. GLYCOCONJUGATES

Structural analysis of proteins has shown that up to half of naturally occurring proteins are subject to post-translational modifications with the vast majority glycosylated. These covalent linkages involve several amino acids and have distinct structrual characteristics. In addition, a large number of lipids have covalently attached carbohydrate, necessary for their biological functions.

The glycoproteins may be classified into two broad categories: *N*-linked and *O*-linked.

A. *N*-Linked Glycoproteins

N-linked glycoproteins have carbohydrate covalently attached to asparagine residues that occur in the sequence Asn-X-Ser/Thr, where X is any residue except proline. This is a necessary but not sufficient key for glycosylation since there are many examples of such sequences that are not glycosylated even when others on the same polypeptide are substituted with sugar. The linking sugar is invariably *N*-acetylglucosamine, which is the terminal saccharide of the attached unit (Fig. 25). The number of saccharides present in *N*-linked structures varies from about 7 to 20 or more; branching is universal with some structures having four separate branches (antennae). All of the saccharides have a common core structure: GlcNAc-GlcNAc-Man_3. The first mannose is β-linked (unusual) and the other two mannoses are attached α-1-3 and α-1-6, thus forming the initial branch point. Unlike protein synthesis wherein the amino acid sequence is controlled by the genetic one, the final structure of saccharides is rarely so conserved. Thus, a given protein with several glycosylation loci is quite likely to have differing saccharide structures, even at the same amino acid site—all of them will still contain the pentasaccharide core noted above. These various glycoforms give rise to a type of heterogeneity that is difficult to characterize completely and may have implications for function.

The biosynthesis of these molecules is also unusual. The saccharide is preassembled, not as the final structure but as a common, 14-sugar, lipid-linked precursor that is transferred *en bloc* to the target asparagine in a cotranslational manner (Fig. 26). This saccharide unit ($GlcNAc_2$-Man_9-Glc_3) is trimmed to a $GlcNAc_2$-Man_5 structure that is then modified by addition of either more mannosyl residues or by several sugars, including GlcNAc, Gal, NANA, and L-fucose. The latter category is generally termed *complex* as opposed to those which contain GlcNAc and Man only (high mannose).

B. *O*-Linked Glycoconjugates

O-linked glycoconjugates have substantial diversity in that the saccharide units may be covalently attached to serine, threonine, tyrosine, hydroxylysine, or hydroxyproline residues. In addition, the type of glycosyl substitution

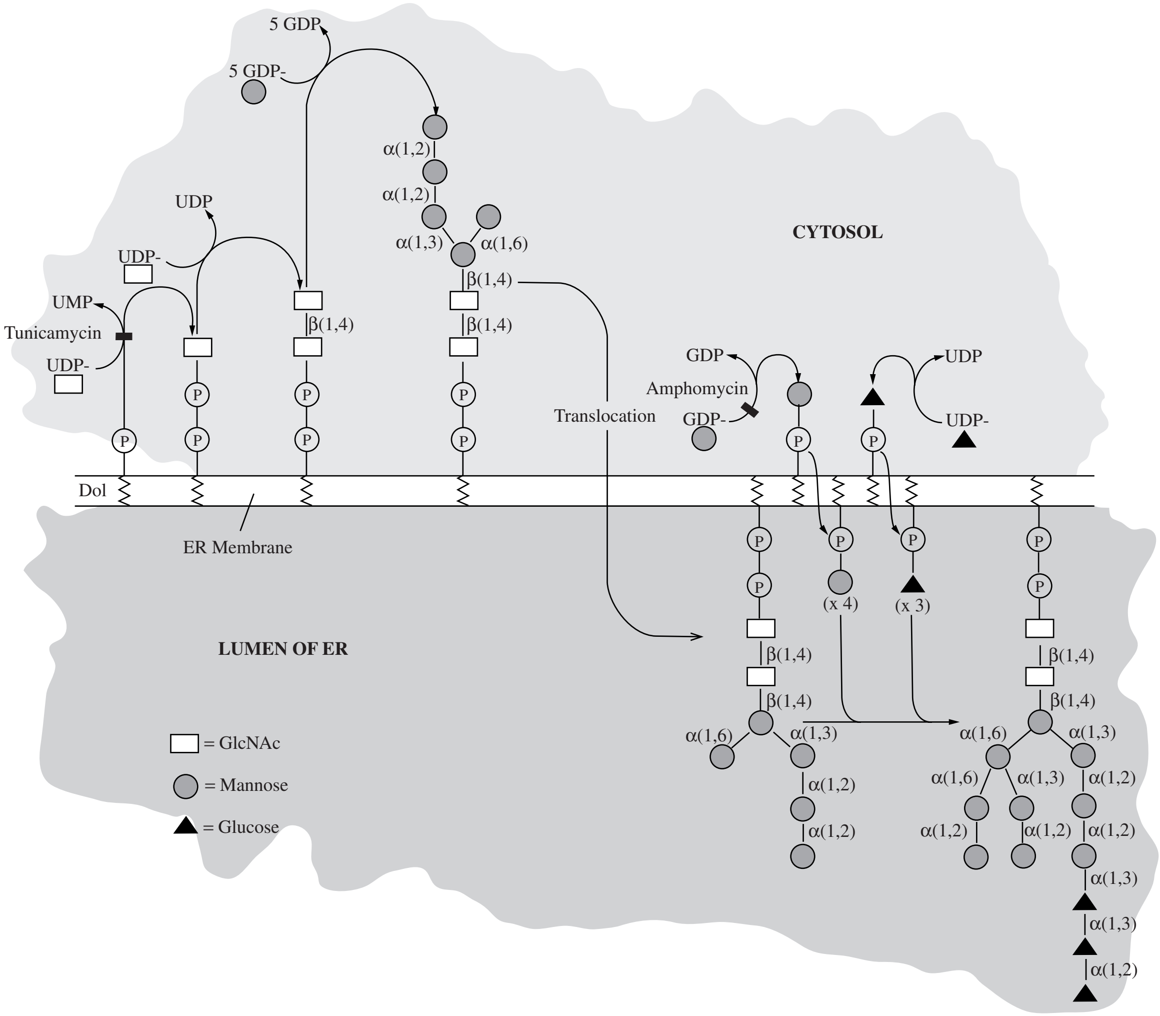


FIGURE 26 Biosynthesis of the 14-sugar, lipid-linked oligosaccharide, the universal precursor for *N*-linked glycosylation.

varies widely, from single sugars to extended polysaccharide chains. The following discussion highlights key features of these types but is not intended to provide full details.

One major category of *O*-linked glycosylation is termed *mucin type*. This is characterized by linkage of the sugar (*N*-acetylgalactosamine in the alpha configuration) to serine or threonine hydroxyl groups (Fig. 27). There is no identifiable consensus amino acid sequence known which targets specific residues to be substituted. The saccharide units range from di- to intermediate size oligosaccharides (up to 10 sugars) and are very diverse. Additional sugars present include galactose, *N*-acetylglucosamine, L-fucose, and sialic acid; some of the saccharide units may be sulfated. Mannose is characteristically absent. These molecules are often found in epithelial secretions; the protein cores may be quite large with a single glycoprotein having an aggregate molecular weight of one million with a hundred or more saccharide units covalently attached.

Glycosylation of tyrosine residues is unusual but a key step in the biosynthesis of glycogen, the major storage glucan of liver and muscle. The core protein, glycogenin, is able to autoglucosylate and attaches a series of glucosyl residues to a single tyrosine in the protein. When the glucose chain has reached four (or more) units (all linked α-1-4), the resulting saccharide moiety is then recognized by glycogen synthase for continuation of glycogen formation. The final polysaccharide may have several thousand glucose residues.

Currently about 20 proteins have been identified as collagens. Criteria for this classification include the presence of a triple helical domain ("collagen helix") and the presence of hydroxyproline and hydroxylysine residues. The latter may also be glycosylated with either a single galactose residue or a disaccharide (glucopyranosyl

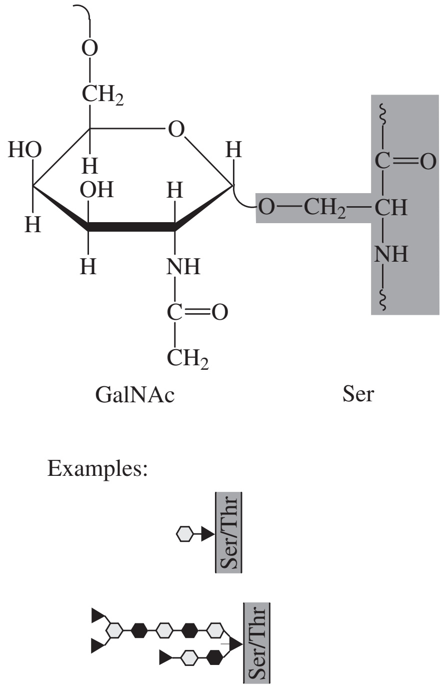


FIGURE 27 Typical *O*-linked, mucin-type oligosaccharide.

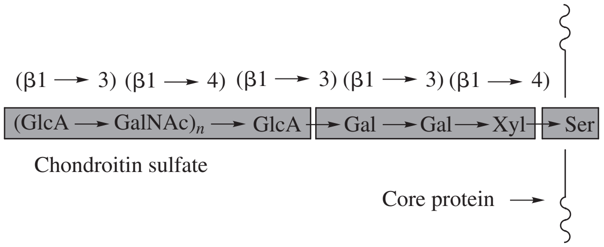


FIGURE 28 Core saccharide of proteoglycans.

1-4 galactosyl-hydroxylysine). The extent of glycosylation varies from as little as 1 per 1000 amino acids to 8 or more. It is suggested that addition of carbohydrate causes local disruption in the collagen helix, thus opening the protein structure into a more mesh-like conformation.

Glycosylation on hydroxyproline residues is rare but has been reported in both plants and fungi. Interestingly, this does not occur in collagen where hydroxyproline is a prominent residue.

An unusual form of *O*-glycosylation has recently been described wherein a single *N*-acetylglucosaminyl residue is attached to either a serine or threonine residue in target proteins. In contrast to other *O*-linked glycoproteins, the entities involved are cytosolic and not secreted. It has been suggested that this modification is reciprocal with phosphorylation, a common form of regulatory substitution.

A major class of *O*-linked glycoconjugates is the proteoglycans, key components of the extracellular matrix of animals. These glycoconjugates are distinguished by the linking sugar (D-xylose, attached to serine), a linear core saccharide (Gal-Gal-Xyl; Fig. 28), and continuation of the saccharide chain as a linear polysaccharide, which contains alternating residues of an amino sugar (*N*-acetylglucosamine or *N*-acetylgalactosamine) and a uronic acid (D-glucuronic or L-iduronic acid) (Fig. 29). The saccharides are generally sulfated (may include *N*-sulfation instead of *N*-acetylation of the glucosaminyl residues) giving rise to chains of considerable structural diversity. Examples include heparin, a natural anticoagulant, and the chondroitin sulfate proteoglycans. In the case of heparin, it has been established that the antithrombin activity resides in a specific pentasaccharide sequence within the structure with a defined pattern of sulfation and sugar components. This type of "information" is known to be present in other complex saccharides and broadens the function of these molecules beyond that of space occupancy and water and electrolyte management. It is interesting to note that the biosynthesis of the iduronosyl moiety in heparin and related heparan sulfate chains occurs at the polymer level by inversion of configuration at C-5 of already incorporated glucuronosyl residues. Extracellular proteoglycans such as those of the chondroitin and dermatan sulfate families are associated with organization of the fibrillar elements of connective tissues (primarily collagens), bone deposition and maintenance of tissue hydration.

Glycolipids represent another diverse class of glycoconjugates. In this case, the saccharides are assembled on a nitrogenous lipid (ceramide) derived from a C-18 amino alcohol (sphingosine) by fatty acylation of the amino group (Fig. 30). The primary hydroxyl group at C-1 is the site of sugar attachment. The saccharides range up to 10 sugars

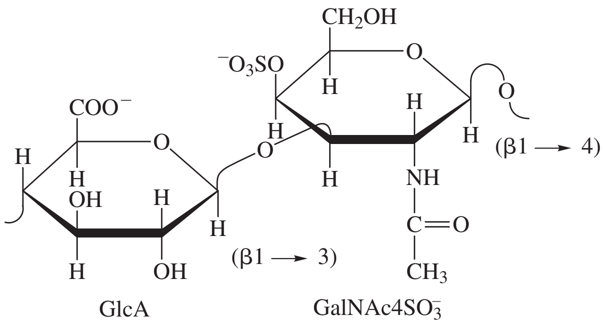


FIGURE 29 Structure of the repeating unit of chondroitin 4-sulfate. Other glycosaminoglycan chains presents in proteoglycans include dermatan sulfate (L-iduronic acid replacing D-glucuronic acid), variants with sulfate in the 6-position, and the heparin–heparan sulfate family, which contains both uronic acids, glucosamine, and both *N*- and *O*-sulfate esters.

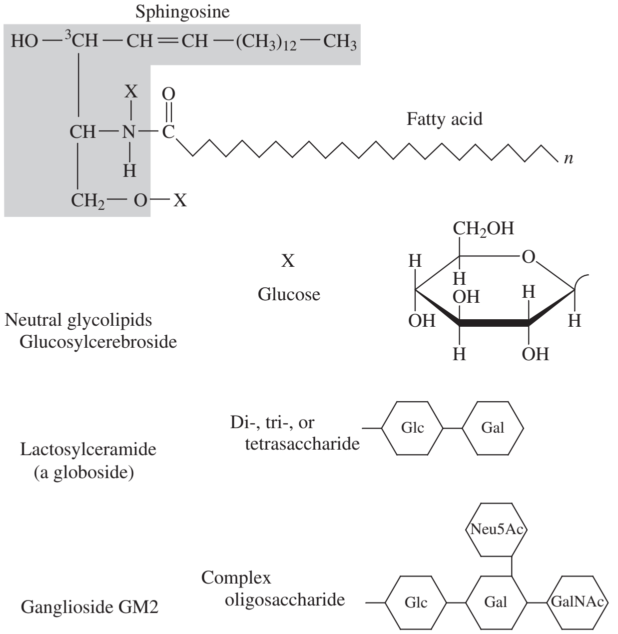


FIGURE 30 Ceramide and glycosphingolipid structures.

and are extremely diverse. They are found on cell surfaces and function as receptors and immunologic determinants. The blood group ABO system is defined by specific sugars present on erythrocyte glycolipids; thus, type A is characterized by alpha-linked *N*-acetylgalactosamine, type B by alpha-linked galactose, etc. Glycosphingolipids that contain sialic acid are termed *gangliosides* (originally isolated from neural tissue) and are involved in development, especially in the nervous system.

V. ANALYTICAL METHODS

Classical methods for the determination of saccharide structures are rarely employed at this time. Thus, methylation and periodate analyses, once the mainstay of carbohydrate work, have been superseded by mass spectrometry and nuclear magnetic resonance (NMR).

Consider the typical problem—that of defining the complete structure of an oligosaccharide (polysaccharide), free or combined. Unlike procedures for analysis of DNA or proteins where automated methods provide rapid and unambiguous sequences, sugar sequences are not readily determined. Needed are linkage positions, configurations, and branching as well as identification of the specific sugars. In some cases, the use of specific enzymes can provide information about the linkage configuration but these do not always work, rarely are quantitative, and often do not discriminate between linkages. The advent of NMR has permitted complete analysis of oligosaccharides as long as sufficient material is available (or spectrometer time). Thus, the linkage configuration for pyranosides is readily apparent from the coupling constant between the protons on carbons-1 and -2 of the sugar in question. The order of sugars and possible branching can be determined

TABLE I Representative Lectins and Their Ligands[a]

Lectin family and lectin	Abbreviation	Ligand(s)
Plant		
Concanavalin A	ConA	Manα1—OCH_3
Griffonia simplicifolia lectin 4	GS4	Lewis b (Le^b) tetrasaccharide
Wheat germ agglutinin	WGA	Neu5Ac(α2 $\rightarrow$ 3)Gal(β1 $\rightarrow$ 4)Glc
		GlcNAc(β1 $\rightarrow$ 4)GlcNAc
Ricin		Gal(β1 $\rightarrow$ 4)Glc
Animal		
Galectin-1		Gal(β1 $\rightarrow$ 4)Glc
Mannose-binding protein A	MBP-A	High-mannose octasaccharide
Viral		
Influenza virus hemagglutinin	HA	Neu5Ac(α2 $\rightarrow$ 6)Gal(β1 $\rightarrow$ 4)Glc
Polyoma virus protein 1	VP1	Neu5Ac(α2 $\rightarrow$ 3)Gal(β1 $\rightarrow$ 4)Glc
Bacterial		
Enterotoxin	LT	Gal
Cholera toxin	CT	GM1 pentasaccharide

Source: Weiss, W. I., and Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65,** 441–473.

[a] More than 100 lectins are known. X-ray diffraction analyses of crystalline lectins suggest a common protein architecture for the saccharide binding sites.

directly using mass spectrometric methods. In this case, however, direct identification of a specific hexose may not be possible since all will have the same molecular weight. Therefore, such studies are generally combined with compositional analysis (acid-catalyzed hydrolysis followed by chromatographic separation) and prior knowledge of the type of saccharide involved.

Some structural data may also be obtained by the use of specific lectins (Table 1). These are a class of carbohydrate-binding proteins, generally of plant origin, that show high specificity for one or another saccharide. This may include linkage configuration and recognition of oligosaccharide.

Information desired for polymeric structures beyond that of components and linkages will generally include molecular weight and solution conformation. Since polysaccharides (including chains found in proteoglycans) are polydisperse, the normal definition of molecular weight (as that of a single chemical entity) does not apply. Ideally, one should determine the mole fraction of each species present as a function of degree of polymerization. Currently, there are no methods for achieving this goal. Rather, physical measurements are made that provide number average (osmotic pressure, suitable only for relatively low molecular weights and limited by solubility), weight average (sedimentation or light scattering), and viscosity average (a mixed function that is also a measure of solution conformation—a measurement of mean end-to-end distance). Solution structures are often helical for repeating linear polysaccharides and heavily dependent on solvent (ionic strength) for polyelectrolytes.

SEE ALSO THE FOLLOWING ARTICLES

Biopolymers • Kinetics (Chemistry) • Pharmaceuticals • Organic Chemistry, Synthesis • Polymers, Structure • Protein Structure • Protein Synthesis • Stereochemistry • Thermodynamics

BIBLIOGRAPHY

Advances in Carbohydrate Chemistry and Biochemistry—yearly review volume published by Academic Press, New York.

Pigman, W., and Horton, D. eds., (1970). "The Carbohydrates, Chemistry and Biochemistry," 4 volumes, Academic Press, New York.

Stick, Robert V. (2001). "Carbohydrates: the Sweet Molecules of Life," Academic Press, New York.

Ion Transport Across Biological Membranes

George P. Hess
Cornell University

I. Relationship Between Transmembrane Inorganic Ion Flux and Transmembrane Potential
II. Mechanism of Transmembrane Inorganic Ion Flux
III. Inorganic Ion Transport and the Perception of Light
IV. Inorganic Ion Transport and Integration of Environmental Information
V. Inorganic Ion Transport and the Rapid Transmission of Electrical Signals over Long Distances (up to ~1 m)
VI. Properties of the Protein (Potassium Channel) that Allows K^+ but Not Na^+ to Cross the Membrane
VII. Inorganic Ion Transport against a Concentration Gradient at Expense of ATP Hydrolysis
VIII. Conclusion and Outlook

GLOSSARY

Adenosine triphosphate (ATP) Upon hydrolysis within a cell, ATP gives rise to adenosine diphosphate and adenosine monophosphate with the liberation of energy (ΔG is about − 12 kcal/mol).

Faraday constant F Has a value of 96,485 coulombs per mole.

Guanosine 3′,5′-cyclic monophosphate (cGMP) Is converted to 5′ guanosine monophosphate by an enzyme, cGMP phosphodiesterase.

Universal gas constant R Has a value of 8.314 joules K^{-1} mol^{-1} (K is degrees Kelvin).

THE MOVEMENT of inorganic ions across biological membranes of animals plays a central role in the perception and integration of, and reaction to, environmental signals by the organism. Examples include vision, the integration and processing of this information (brain function), and the reaction of the organism to this information, for instance, muscle contraction (Fig. 1). Cells have the ability to hydrolyze adenosine triphosphate (ATP). The energy thus released is used to transport sodium and potassium ions across the cell membrane against a concentration gradient. This process establishes the transmembrane voltage (V_m) of the cell membrane. The transmembrane voltage is perturbed by the movement of inorganic ions (generally Na^+, K^+, Cl^-, and Ca^{2+}) along the concentration gradient and voltage difference across the cell membrane that occurs in signal transmission between cells. There are many different transmembrane channel-forming proteins, which are activated by (1) a concentration gradient of inorganic ions across the membrane, (2) the transmembrane voltage,

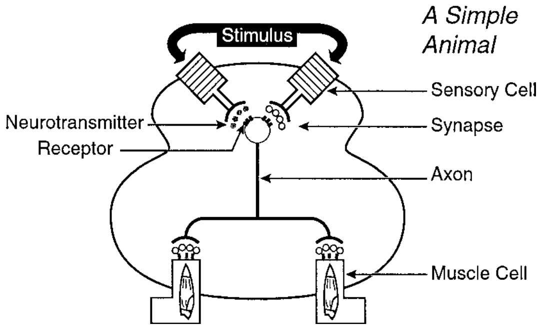


FIGURE 1 An environmental stimulus, for instance, light, activates a protein-mediated reaction in the eye, leading to the transmembrane flux of inorganic ions, a change in the transmembrane voltage, and neurotransmitter release. The neurotransmitter diffuses across a junction between the nerve terminal of the axon and the cell body of the adjacent cell about 20–40 nm in length, called synapse. The neurotransmitter binds to receptors in the membrane of the postsynaptic cell. Excitatory neurotransmitters (○) activate receptors that form cation-specific transmembrane channels. Inhibitory neurotransmitters (●) activate receptors that form anion-specific transmembrane channels. Once the transmembrane voltage of the cell is changed by a critical amplitude and sign (by ∼+20 mV), an all-or-none process occurs. Transmembrane Na^+ and K^+ channels in the axonal membrane open transiently, resulting in an electrical signal that travels down the axon and neurotransmitter is again secreted. This process repeats itself and is terminated when the neurotransmitter is released adjacent to receptors on the surface of muscle cells. In the case of muscle cells, the receptor is the muscle nicotinic acetylcholine receptor and the neurotransmitter acetylcholine. The voltage change in the muscle cell membrane initiates muscle contraction. (From Hess, G. P., and Grewer, C. (1998). "Methods in Enzymology" (G. Marriott, ed.), Vol. 291, pp. 443–474, Academic Press, New York.) The resulting flow of inorganic ions through the membrane of the muscle cell results in a change of its transmembrane voltage V_m and muscle contraction.

(3) the binding of specific ligands to a channel-forming protein. (4) Some proteins use the energy liberated by the hydrolysis of ATP to transport inorganic ions against a concentration gradient. Only one example of each of these various proteins will be mentioned. In each case, the protein chosen is the one about which we have the most information. The proteins that facilitate inorganic ion transport across biological membranes are discussed in an order that illustrates their function in the life of an organism.

I. RELATIONSHIP BETWEEN TRANSMEMBRANE INORGANIC ION FLUX AND TRANSMEMBRANE POTENTIAL

Membranes surround the individual cells of animals and organelles within the cell. They are composed of lipids and proteins. Specific proteins are responsible for the transport of specific inorganic ions across the membrane. Invariably, this transport of inorganic ions across the cell membrane is accompanied by changes in the transmembrane voltage. The equilibrium transmembrane potential for a specific inorganic ion, for instance K^+, is given by the Nernst equation:

$$E_{K^+} = \frac{RT}{ZF} \ln \frac{[K^+]_o}{[K^+]_i}. \tag{1}$$

R, T, and F are the molar gas constant, absolute temperature, and Faraday constant, respectively. Z represents the valence of the inorganic ion and the subscripts o and i indicate whether the ion is outside or inside the cell membrane. The transmembrane potential of neurons is around −60 mV. In the nervous system, changes in the transmembrane potential due to a change in the flux rate of inorganic ions can be propagated rapidly and over distances as long as several feet *via* the axon, a long projection of many nerve cells (Fig. 1). At the axonal terminal, the voltage change initiates a process leading to the flux of calcium ions into the nerve terminal. This results in the secretion of chemical signals, neurotransmitters, which bind to membrane-bound proteins, neurotransmitter receptors, on adjacent cells. Upon binding specific neurotransmitters, the receptors transiently open transmembrane channels. The channels are permeable to Na^+, K^+, or Cl^-, depending on the receptor. The resulting changes in the transmembrane voltage may lead to propagation of a signal to an adjacent cell. Thus, this interplay between chemical reactions and transmembrane voltage changes plays a decisive role in the rapid communication between nerve (and nerve and muscle) cells and in nervous system function.

In 1890, Max Planck derived the relationship between the rate of movement of inorganic cations and anions across a porous barrier and the resulting electric field. If one assumes a constant electric field and constant inorganic ion concentration, the Planck equation is easily integrated and can be used to estimate the transmembrane voltage change, V_m, that results from the flow of inorganic ions across cell membranes. The resulting Goldman equation is

$$V_m = \frac{RT}{F} \ln \frac{P_K(K^+)_o + P_{Na}(Na^+)_o + P_{Cl}[Cl^-]_i}{P_K(K^+)_i + P_{Na}(Na^+)_i + P_{Cl}[Cl^-]_o}. \tag{2}$$

P_K, P_{Na}, and P_{Cl^-} represent the permeability coefficient of the membrane for K^+, Na^+, and Cl^-, respectively. $[K^+]$, $[Na^+]$, $[Cl^-]$ represent the molar concentrations of the ions, and the subscript o or i indicates whether the ions are outside or inside the cell membrane. As usual, R, T, and F represent the molar gas constant, the absolute temperature, and the Faraday constant respectively.

How is the rate of ion movement through a protein-formed channel across a cell membrane related to the

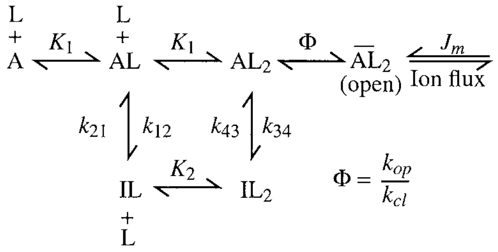


FIGURE 2 Minimum mechanism to account for the rates of a neurotransmitter (acetylcholine) receptor-mediated cation translocation and for receptor inactivation and reactivation as a function of acetylcholine concentration. The active (A) and inactive (I) forms of the receptor bind neurotransmitter (L) in rapidly achieved equilibria denoted by the microscopic equilibrium constants (K). Active receptor with two bound ligand molecules (AL_2) converts rapidly (1 to 2 msec) to an open channel ($\overline{AL}_2$) with an equilibrium constant for channel opening ($1/\Phi \cdot \Phi = k_{cl}/k_{op}$ where k_{op} and k_{cl} are the rate constants for channel opening and closing respectively). $\overline{AL}_2$ permits the movement of inorganic Na^+ and K^+ ions through the membrane, where J_m, is the observed rate constant for the flux of inorganic ions (Na^+, K^+) through the open receptor-formed transmembrane channel (see Eq. 3). In the continued presence of neurotransmitter, the receptors reversibly form inactive forms I in the 10- to 200-msec time region, depending on the receptor and the concentration of neurotransmitter. This process is called receptor desensitization. (Reproduced with permission from Cash, D. J., Aoshima, H., and Hess, G. P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3381–3322.)

transmembrane voltage? In the nervous system, signal transmission is regulated by the binding of chemical signals, neurotransmitters, to membrane-bound proteins, called receptors. Commonly, when two molecules of a neurotransmitter have bound to the receptor, the protein forms a transmembrane channel that remains open for a few milliseconds, allowing the receptor-specific passage of sodium, potassium, or chloride ions. The chemical reaction for many neurotransmitter receptors can be written as shown in Fig. 2; in this case the kinetic mechanism of the nicotinic acetylcholine receptor is used as an example. This receptor plays an important role in signal transmission between nerve cells in the brain and between nerve and muscle cells (Fig. 1).

The specific reaction rate for the transmembrane flux of inorganic cations controlled by the nicotinic acetylcholine receptor, $\bar{J}_m$, has a value of about 5×10^7 M^{-1} sec^{-1} at 14°C. The relationship between the permeability coefficient P for a specific inorganic ion $M^{\pm}$ (Eq. 2) and the specific reaction rate $\bar{J}_m$ is given by:

$$P_{M^{\pm}} = \bar{J}_{M^{\pm}} R_0 (\overline{AL}_2). \qquad (3)$$

R_0 represents the moles of specific receptors in the cell membrane and ($\overline{AL}_2$) the fraction of the receptors that are in the open-channel form. Equations 2 and 3, therefore, establish the important relationship between the receptor-controlled movement of inorganic ions through the cell membrane and the resulting change in transmembrane voltage, V_m.

II. MECHANISM OF TRANSMEMBRANE INORGANIC ION FLUX

Here, the emphasis is on kinetic techniques used to obtain the information needed to understand the mechanism of protein-mediated reactions that allow the transport of inorganic ions across biological membranes. A combination of structural, thermodynamic, and kinetic information is required to achieve this understanding. The use of X-ray crystallography, NMR measurements, and electron microscopy, to obtain structural information about proteins is described in detail elsewhere in this *Encyclopedia.*

In 1976, Neher and Sakmann developed the single-channel current-recording technique. It is simple, convenient, and widely used for measuring the properties of a single, open receptor channel, such as its conductance, lifetime, and ion specificity. In brief, a glass pipet with an internal diameter of 1 to 2 μ is attached to the surface of a cell membrane (Fig. 3A, left). Gentle suction is applied to isolate a small membrane patch within the glass pipet from the rest of the membrane (Fig. 3A, right). A silver chloride wire running from near the tip of the glass pipet to electrical recording equipment allows one to record the current flowing through single receptor-formed channels within the membrane patch (Fig. 3B). In the illustration, the deviation of the current from the baseline represents the current flowing through a single nicotinic acetylcholine receptor-channel in the presence of 20 μM acetylcholine, which activates this channel. The Gaussian distribution of the current amplitude gave a peak centered at 3 pA. The exponential distribution of the time the channel remains open (the lifetime distribution) gave a value of 2.4 msec for the mean lifetime of the open channel, τ_{op}. This lifetime is a measure of the rate constant for channel closing $1/\tau_{op} = k_{cl}$ (Fig. 2). Additionally, the technique allows one to determine conveniently the ion specificity of the channel from measurements of the current amplitude.

In the case of transmembrane channels that open upon binding a specific ligand, how does one determine the other parameters of the channel-opening reactions? These are the dissociation constant of the neurotransmitter from the site controlling channel opening, the rate constant for channel opening k_{op} and, therefore, the equilibrium constant for channel opening, $1/\Phi = k_{op}/k_{cl}$. In principle, many of these constants can be determined from the lifetime of the closed state, the time interval between the openings of single receptor-channels (Fig. 3B). From the mechanism in Fig. 2, a plot of the number of closures observed

FIGURE 3 The single-channel current-recording technique. (A) The tip of a borosilicate glass pipet, with a tip opening of 1–2 μm is pressed against the membrane of a frog muscle cell (left). A slight negative pressure (20–30 cm H_2O) is applied to the pipet for several seconds to form the seal between the membrane and the pipet (right). (Reproduced from O. Hamill *et al.* (1995). *In* "Single-Channel Recording," 2nd edition (B. Sakmann and E. Neher, eds.), p. 663, Plenum Press, New York.) (B) A typical current trace recorded using the single-channel technique, a rat myoball cell containing nicotinic acetylcholine receptors, and 20-μM acetylcholine (pH 7.2, 22°C, and $V_m = -80$ mV). (Reproduced from F. Sigworth (1983). *In* "Single-Channel Recording," first edition (B. Sakmann and E. Neher, eds.), Plenum Press, New York.)

within a definite time interval *versus* the time the channel was closed is expected to give a three-exponential distribution. From this distribution, the three different lifetimes, reflecting the constants to be determined, can be calculated. This evaluation requires many measurements to be made, which take time, and it is restricted to measurements made at low concentrations of neurotransmitters. At higher concentrations of neurotransmitter, the receptor becomes inactive, desensitized (in the millisecond time region) (Fig. 2) and the signal to be measured disappears. Additionally, we now know that many receptors on the cell surface exist in two forms, which desensitize with different rates.

The desired, and missing, information that supplements results obtained with the single-channel current–recording technique can now be obtained by using a transient kinetic method with a microsecond time resolution, the laser-pulse photolysis (LaPP) technique. The usual rapid kinetic techniques that are suitable for investigating small molecules in solution had to be modified for use with membrane-bound proteins. The time resolution for equilibrating ligands in solution with membrane-bound proteins is less than might be expected. This is because a layer of water molecules (the diffusion layer) covers the membrane containing the proteins on the surface of relatively large objects like cells, or even membrane patches with diameters in the micrometer range. Ligands in the solution surrounding the membrane-bound receptors must diffuse through the diffusion layer and this process may become rate limiting. The steps needed to overcome this problem are illustrated in Fig. 4.

Photolabile precursors of neurotransmitters ("caged" neurotransmitters) that are biologically inactive have been developed. A photolabile precursor of carbamoylcholine, a stable analog of acetylcholine that activates the nicotinic acetylcholine receptor, is shown in the inset to Fig. 4. Photolabile precursors of all the major neurotransmitters are now available. This photolabile precursor of carbamoylcholine ([*N*-(α-carboxy-2-nitrobenzyl)-carbamoylcholine] is equilibrated with nicotine acetylcholine receptors on the surface of a cell (Fig. 4A). At zero time, the compound is photolyzed, using a laser, by a single pulse within about 100 μsec to give carbamoylcholine and a biologically inert side-product, a 2-nitroso-α-keto carboxylic acid (Fig. 4 reaction). An optical fiber carries the light beam to the cell, which is attached to a current-recording electrode (Fig. 4A). The technique for recording the current from all the receptors on the cell surface with high precision uses the same equipment as is used in the single-channel current-recording technique (Fig. 3). The increase in current that results when carbamoylcholine is liberated on the cell surface, due to the photolysis of caged carbamoylcholine is shown in Fig. 4B. The current is due the opening of receptor-channels on the cell surface and the flow of inorganic ions through them. In a different and slower time zone, the current then decreases due to receptor desensitization. In experiments with different neurotransmitter [glutamate, serotonin, γ-aminobutyric acid (GABA), and glycine] receptors, conditions could be obtained in which the rise of the current follows a single exponential rate law. The observed rate constant for the rise time, k_{obs}, is related to the rate constants for channel opening (k_{op}) and closing (k_{cl}), the concentration of the ligand L that activates the transmembrane channel, and the dissociation constant of L, that is, K_1 (Fig. 2):

$$k_{obs} = k_{cl} + k_{op}[L/(L + K_1)]^2. \tag{4}$$

The relationship between k_{obs} and the concentration of neurotransmitter is given in Fig. 4C. The slope of the line gives the value of the rate constant for channel opening (k_{op}) and the intercept on the ordinate gives the rate constant for channel closing (k_{cl}).

This section outlined some approaches used to study the mechanism of proteins that transport inorganic ions across biological membranes. In the next section the properties of some individual proteins will be discussed.

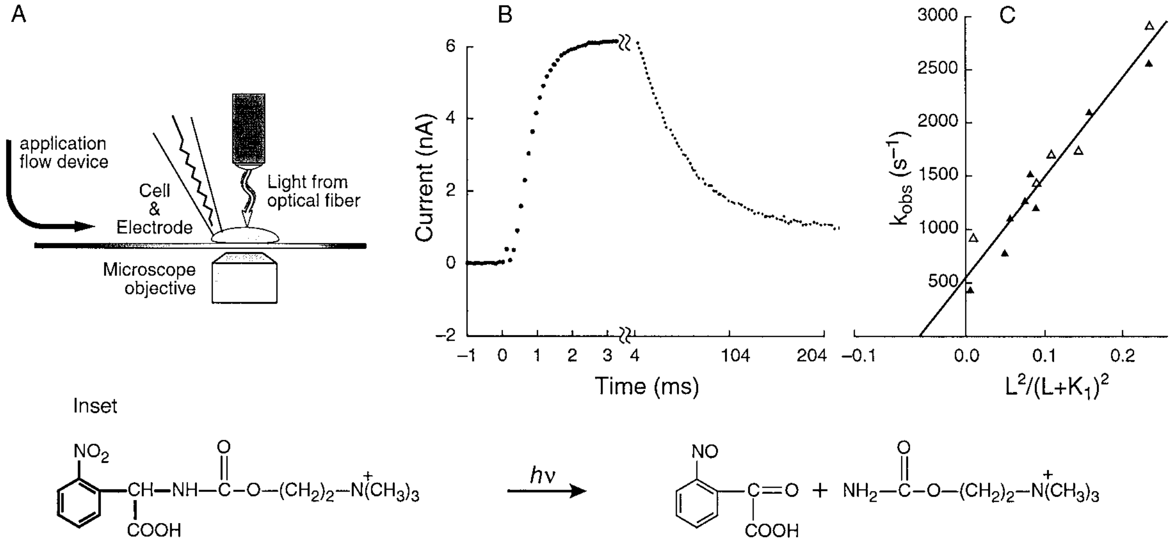


FIGURE 4 A BC_3H1 cell (~20-μm diameter) containing nicotinic acetylcholine receptors, attached to an electrode for whole-cell current recording, is equilibrated with caged carbamoylcholine. A cell-flow device was used to equilibrate the cell surface with ligands in a solution flowing over the cell from the left. (A) A Candela SLL500 dye laser is used. Rhodamine 640 or sulforhodamine 640 laser dye, together with a second harmonic generator, produces wavelengths of 328 and 318 nm, respectively. The laser beam is introduced from an optical fiber of 200-μm diameter. The fiber is adjusted to be ~400 μm away from the cells so that the area illuminated around the cell has a diameter of 300 to 400 μm. The energy of the laser pulse emerging from the fiber is ~500 μJ and the pulse length is 600 nsec. By projecting visible light through the optical fiber, the cell is illuminated and the fiber properly positioned. Inset: Photolysis at 328 nm (using the Candela dye laser) of caged carbamoylcholine liberates a 2-nitroso-α-ketocarboxylic acid and carbamoylcholine, a stable and well-characterized analog of acetylcholine. The wavelength of 328 nm was chosen to avoid cell damage at lower wavelengths and too low a product yield at higher wavelengths. Current is recorded in the whole-cell configuration. (B) Whole-cell current induced by 200 μM released carbamoylcholine at pH 7.4, 22–23°C, and −60 mV. The points represent the digitized current data; the parameters of the heavy solid line were calculated from a first-order plot of the rise time and an observed first-order rate constant of 2140 sec^{-1}. (C) Determination of k_{op}, k_{cl}, and K_1 for the opening of nicotinic acetylcholine receptor channels in BC_3H1 cells at pH 7.4, 23°C and −60 mV. The values of k_{obs} determined from experiments shown in B are plotted according to $k_{obs} = k_{cl} + k_{op}\, L^2/(L + K_1)^2$. The values of k_{op}, k_{cl}, and K_1 are 12,000 sec^{-1}, 500 sec^{-1}, and 210 μM, respectively. (Reproduced with permission from Hess, G. P., and Grewer, C. (1998). *In* "Methods in Enzymology," (G. Marriott, Ed.), Academic Press, New York.

III. INORGANIC ION TRANSPORT AND THE PERCEPTION OF LIGHT

We shall now follow a cascade that consists of the perception of an environmental signal, the movement of ions across membranes and the resulting changes in transmembrane voltage, and the reaction of the organism to the signal (Fig. 1). Because we know most about the sensory cells of the eye, this system is used as an example, and particularly the role of rhodopsin.

Opsin, which loops back and forth across the cell membrane, and rhodopsin are transmembrane proteins. Unwin and Henderson used electron microscopy to determine the structure of bacterial rhodopsin at 7 Å resolution. More recently, the crystal structure of bovine rhodopsin at 2.8 Å resolution was solved by Palczewski and colleagues.

Perception of light is initiated by a photochemical reaction in the retinal cells of the eye, the *cis–trans* isomerization of 11-*cis*-retinal (Fig. 5A). 11-*Cis*-retinal is attached by a Schiff-base linkage between its aldehyde group and the ϵ-amino group of a lysine residue in the protein opsin, thus forming rhodopsin. 11-*Cis*-retinal absorbs light in the visible wavelength region, with an absorption maximum at 500 nm and a molar extinction coefficient of 40,000 cm^{-1} M^{-1}. The absorption of a single photon of light by 11-*cis*-retinal, and the subsequent conversion to all-*trans*-retinal, initiates a series of reactions. These include a conformational change of rhodopsin, and lead to a change in the rate at which sodium ions cross the membrane (see the next paragraph). This results in a change of the voltage across the cell membrane. The same conformational change of rhodopsin, initiated by the absorption of

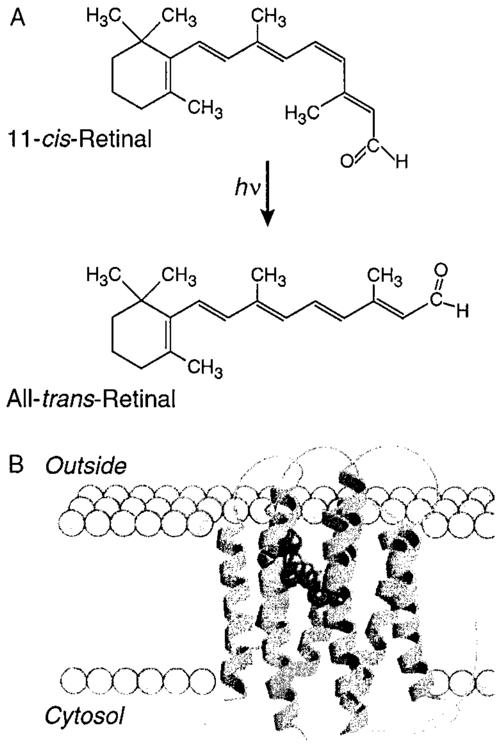


FIGURE 5 (A) Light-induced transformation of 11-*cis*-retinal to all-*trans*-retinal in visual pigments. 11-*Cis*-retinal is attached through a Schiff-base linkage of lysine 256 in rhodopsin. (B) Rhodopsin has a molecular weight of 40,000. Seven transmembrane helices are embedded in the cell membranes. 11-*Cis*-retinal lies near the center of the lipid membrane. The structure is based on the three-dimensional reconstruction of electron microscope images by Henderson and Unwin. (Reproduced with permission from Nelson, D. L., and Cox, M. M. p. 460. Figs. 13–22. "Lehninger Principles of Biochemistry" (2000). 3rd edition, Worth Publishers p. 460.)

light, leads to the release of all-*trans*-retinal. The all-*trans*-retinal is transformed to all-*trans*-retinol (Vitamin A) by pigment epithelia cells in the eye. It is a precursor in the synthesis of 11-*cis*-retinal, which is then transported back to cells containing opsin and reattached to the ϵ-amino group of lysine 296 of opsin.

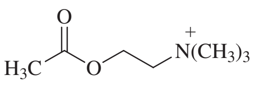
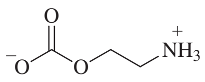
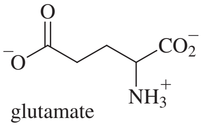
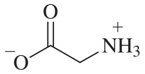
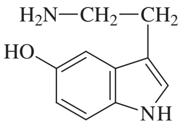
The steps involved in the absorption of light by rhodopsin that lead to changes in the flux of sodium ions across the cell membrane, and to signal transmission, can be summarized as follows. (1) Light is absorbed. (2) Subsequently, the isomerization of retinal results in (3) a conformational change of rhodopsin; leading to (4) the activation of the enzyme cyclic-guanosine monophosphate (cGMP) phosphodiesterase. This results in (5) the hydrolysis of cyclic guanosine monophosphate (cGMP) to 5′ guanosine monophosphate (5′-GMP). (6) cGMP activates a protein that forms Na^+-specific channels in the retinal cell membrane and maintains the transmembrane potential V_m at about −40 mV. (7) Lowering the concentration of cGMP in the cell lowers the concentration of open Na^+-conducting channels, which to remain open must have cGMP bound to them. (8) The closing of cGMP-activated transmembrane channels decreases the flux rate of Na^+ into the cell and, therefore, changes the voltage across the membrane. (9) The changes in V_m result in the release by the retinal cell of chemical signals (neurotransmitters) adjacent to another cell. (10) The neurotransmitters bind to receptors on an adjacent cell that transiently form transmembrane channels, allowing cations or anions, depending on the receptor, to move through the cell membrane.

IV. INORGANIC ION TRANSPORT AND INTEGRATION OF ENVIRONMENTAL INFORMATION

In the previous section, the transport of ions across the membrane was initiated by an environmental signal, namely, light. This resulted in a change in the transmembrane voltage and a subsequent influx of calcium ions into the nerve terminal of a sensory cell (Fig. 1), resulting in the release of a chemical signal. In general, these chemical signals (neurotransmitters) are released from the nerve terminals of one cell and diffuse across a gap between cells called a synapse (Fig. 1). In the membrane of the adjacent cell, about 20–40 nm removed from the nerve terminal, the neurotransmitters bind to transmembrane neurotransmitter receptors. On binding the neurotransmitter, these transmembrane proteins, transiently (a few milliseconds) open a transmembrane channel (Fig. 1). These channels are specific for inorganic cations, sodium, or potassium, and some are also specific for calcium or chloride ions, depending on the receptor. If the resulting change in the transmembrane voltage, V_m, is of appropriate sign and magnitude ($\sim +20$ mV) the voltage change will be propagated along the axon of the cell (Fig. 1), by a process in which sodium and potassium ions move across the membrane. This is described in detail at the end of this section.

The major neurotransmitters are listed in Table I. Typical excitatory neurotransmitters are acetylcholine, glutamate, and serotonin. They bind to receptors that are named in the same way, for instance, the acetylcholine receptor, glutamate receptor, etc. These receptors on binding their specific neurotransmitter form transmembrane channels specific for the inorganic sodium and potassium cations. They are called excitatory receptors because they shift the transmembrane potential of the cell membrane to more positive values. A change in V_m of about +20 mV is required for the electrical signal to be transmitted to the nerve terminal where the release of a neurotransmitter is elicited. Typical inhibitory neurotransmitters are glycine

TABLE I Major Neurotransmitters

Excitatory	Inhibitory
 acetylcholine	 GABA
 glutamate	 glycine
 serotonin	

and γ-aminobutyric acid (GABA). They bind to receptors that form transmembrane channels that are specific for chloride ions. They are called inhibitory receptors because they shift the transmembrane potential to more negative values and counteract the action of excitatory receptors.

The neurotransmitter receptors are believed to belong to one large family of transmembrane, channel-forming proteins. A three-dimensional model of the nicotinic acetylcholine receptor from the electric organ of a fish, *Torpedo* species, has been proposed by Nigel Unwin based on electron diffraction images. Figure 6 gives a side view

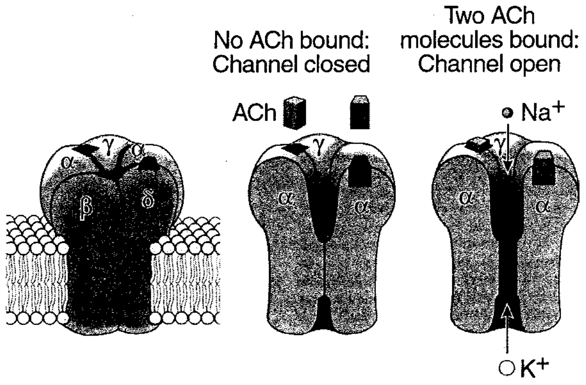


FIGURE 6 Three-dimensional model of the nicotinic acetylcholine-gated ion channel. The structure is based on a three-dimensional reconstruction of electron microscope images by Unwin and his colleagues. The receptor-channel complex consists of five subunits, all of which contribute to forming the pore. When two molecules of acetylcholine bind to portions of the α-subunits exposed to the membrane surface, the receptor-channel changes conformation. This opens a pore in the portion of the channel embedded in the lipid bilayer, and both K^+ and Na^+ flow through the open channel down their electrochemical gradients. (Reproduced with permission from Kandel, E. R., Schwartz, J. H., and Jessel, T. M. (2000). "Principles of Neuronal Science," 4th edition. McGraw Hill, New York.)

of the receptor spanning the membrane. The receptor consists of 5 subunits, each with a molecular weight of about 50 kiloDaltons. Together the subunits form a funnel-shaped structure, projecting about 60 A above the plane of the membrane. The opening of the channel is about 50 A in diameter and narrows to a channel through the membrane about 30 A in length and 8 A wide.

What is the relationship between the neurotransmitter concentration, the fraction of receptor-channels that open, the conductance of the channel, and the length of time the channels remain open (Fig. 2)? If we can answer these questions, and can also determine the concentration of receptor sites exposed to the neurotransmitter, we can calculate the change in transmembrane voltage (Eqs. 1–3) and predict whether or not a signal will be transmitted. The techniques for elucidating the mechanism of the reactions that allow one to determine the rate of transmembrane flux of inorganic ions and, therefore, the change in V_m have been discussed in Section I. It is important to mention that each cell is contacted by as many as 1000 different projections from other cells and each of these projections may release a different neurotransmitter from the nerve terminal. The mammalian brain contains an estimated 10^{12} cells; its information content is believed to exceed, by far, that of the largest supercomputers.

It appears that the techniques are now at hand to (a) investigate the kinetic mechanism of the reactions by which neurotransmitter receptors determine the rate of flow of inorganic ions across the membrane, and (b) determine whether or not a signal will be transmitted to another cell. The chemical mechanism of receptor-mediated transport of inorganic ions across the membrane of neurons (nerve cells) by a few neurotransmitter receptors has been investigated. The mechanisms by which the many receptor isoforms facilitate the flow of inorganic ions across the cell membrane are still unknown. Also virtually unknown are the effects of diseases, causing mutations of receptors, and the influence of hundreds of drugs including anticonvulsants, antidepressants, anesthetics, and abused drugs, on the mechanisms.

V. INORGANIC ION TRANSPORT AND THE RAPID TRANSMISSION OF ELECTRICAL SIGNALS OVER LONG DISTANCES (UP TO ~1 m)

The movement of inorganic ions across a cell membrane that initiates an electrical signal is only one step in signal transmission in organisms. The rapid transmission of this signal, over distances of up to 1 m, by the axon of cells is discussed next.

How is the signal that was initiated by neurotransmitter-mediated reactions propagated? In general, cell membranes are more permeable to potassium ions, which are at a higher concentration inside the cell membrane than outside and bound to immobile cations, mainly amino acids and proteins. The outflow of potassium ions along their concentration gradient is counteracted by the transmembrane voltage that is created by this outflow. The equilibrium potential for potassium ions, for instance, E_{K^+}, is given by the Nernst equation (Eq. 1). The resting transmembrane potential of neurons is around −60 mV.

When the flux of ions through neurotransmitter receptor channels results in a change (of ∼20 mV) in the transmembrane voltage to a more positive V_m value, the electrical signal is propagated along the axon of the cell within 1 msec. This occurs in the following way. At a critical value of V_m, specific voltage-dependent Na^+ transmembrane channels in the axonal membrane open, allowing sodium ions to flow inside the axon (Fig. 1). As the inward flow of sodium ions changes V_m to even more positive values, the Na^+-specific channels are inactivated. Transmembrane K^+ channels open, changing V_m again to more negative values. This in turn leads again to the opening of voltage-dependent Na^+-specific channels. These changes in transmembrane voltage, called action potentials, are propagated along the axon to the nerve terminal adjacent to another cell. Axons range from 0.1 mm to over 1 m in length and can convey electrical signals within 1 msec. When the signal arrives at the nerve terminal, Ca^{2+}-specific transmembrane channels open and the influx of calcium ions leads to the secretion of neurotransmitter. Typically, the neurotransmitter diffuses across the synaptic cleft, over a distance of 20–40 nm, and binds to the neurotransmitter receptors on the adjacent cell. Thus the signal is transmitted between the $\sim 10^{12}$ cells of the mammalian nervous system.

VI. PROPERTIES OF THE PROTEIN (POTASSIUM CHANNEL) THAT ALLOWS K^+ BUT NOT Na^+ TO CROSS THE MEMBRANE

Recently, we have learned some of the properties of one channel that plays a central role in rapid signal transmission. The K^+ channel from bacteria was crystallized, after a cytoplasmic tail of 33 residues was removed, and MacKinnon and colleagues have determined its structure at a resolution of 3.2 A (Fig. 7). This work represents the first high-resolution, X-ray diffraction study of an ion-selective channel. Although the structure was obtained with the bacterial channel, K^+ channels with similar sequences and properties are also present in other organisms. The bacterial K^+ channel contains 158 amino acid residues. Four subunits are arranged around a central axis to form the channel. The K^+ channel has two transmembrane helices.

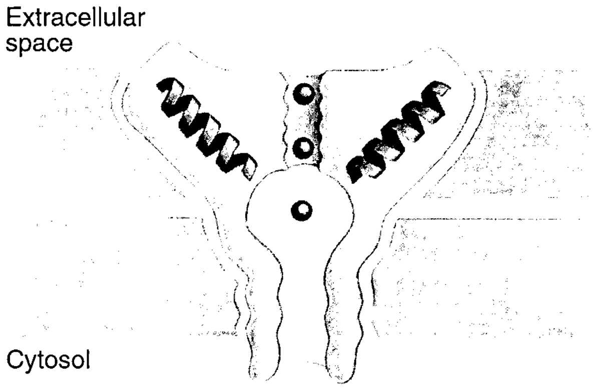


FIGURE 7 The selectivity filter of the potassium channel based on the X-ray crystallographic structure determination by MacKinnon and colleagues. The potassium channel is tetrameric with a hole in the middle that forms the ion pore. Each subunit forms two transmembrane helices, the inner and the outer helix. The pore helix and loop regions build up the ion pore in combination with the inner helix. The black spheres in the middle of the channel represent potassium ions. (Reproduced with permission from Branden, C., and Tooze, J. (1999). Fig. 12.11, p. 233. *In* "Introduction to Protein Structure," 2nd edition, Garland Publishing, New York.)

The structure gives an important indication as to how the channel allows the larger potassium ions (radius 1.35 Å) to move through the channel at 10^8 ions sec^{-1}, which approaches the diffusion-limited rate, but essentially prevents smaller sodium ions (radius 0.95 A) to pass. It has been estimated that K^+ ions move through this channel at least 10,000 times faster than sodium ions. The ion pore is about 45 Å long. Three cation-binding sites have been identified in it, two within a selectivity filter and separated by about 7.5 Å and one in the cavity of the pore (Fig. 7). Both the concentration gradient and the electromotive force provide the driving force moving the ions through the channel. To pass through the channel, the inorganic ions have to pass through a selectivity filter. The theory is that ions that enter the selectivity filter must be dehydrated and that interaction of the dehydrated K^+ with carbonyl oxygens from amino-acid residues inside the filter compensates for the dehydration. The distances between the carboxyl oxygens of the protein and two potassium ions in the filter are optimal to compensate for the cost of dehydration of potassium ions, but they would not be optimal for the dehydration of sodium ions.

VII. INORGANIC ION TRANSPORT AGAINST A CONCENTRATION GRADIENT AT EXPENSE OF ATP HYDROLYSIS

Most cells have a high concentration of K^+ and a low concentration of Na^+ inside the cell membrane relative to the concentration of these ions bathing the cell. As discussed, the rapid flow of these ions along their concentration gradient across the cell membrane is used in signal transduction between the cells of an organism, and plays an important role in life. To maintain the concentration gradient of sodium and potassium ions and thus the resting membrane potential of the cells, energy is needed. In 1979, Skou and Norby discovered an enzyme that reestablishes the original concentration gradient. It is called the Na^+–K^+ ATPase where ATP stands for adenosine triphosphate, an energy source. Sodium ions are moved from inside of the cell to the outside and potassium ions are moved in the opposite direction. With each cycle of the enzyme, one molecule of ATP is hydrolyzed and two potassium ions are moved into the cell and three sodium ions moved out.

Na^+–K^+ ATPase is a membrane protein with two subunits spanning the membrane. The current hypothesis is that the enzyme reacts with ATP to give a phosphorylated enzyme and ADP (adenosine diphosphate). The conversion of ATP to ADP (adenosine diphosphate) and P_i (inorganic phosphate has been formulated as follows:

$$\text{ATP} + \text{Enzyme} \rightarrow \text{ADP} + \text{P-Enzyme (phosphorylated enzyme)}$$

$$\text{P-Enzyme} + H_2O \rightarrow \text{Enzyme} + P_i$$

The resulting enzyme has a high affinity for K^+ and a low affinity for Na^+. Hydrolysis of the phosphorylated enzyme results in the liberation of inorganic phosphate and the regeneration of the enzyme form with high affinity for Na^+ and low affinity for K^+. The net result is the movement of two K^+ ions into the cell and three Na^+ ions out of the cell for each ATP hydrolyzed. About 25% of the energy consumption of a human at rest is used to maintain the resting concentration of sodium and potassium ions in cells.

VIII. CONCLUSION AND OUTLOOK

It should be mentioned that examples of two types of ion channels have been given. (1) So-called slow channels involve second messengers. The examples given here are the light-activated channels that are opened by cGMP and the transmembrane membrane ion transport that requires the hydrolysis of ATP. (2) Fast activated channels. For activation, these just require the binding of a ligand to the channel protein or a change in transmembrane voltage. The examples given are the neurotransmitter-activated channels and the voltage-activated K^+ channel.

The development of techniques using crystallography, NMR, electron diffraction, and molecular biology to produce specific proteins in large amounts to determine the structure of transmembrane channels formed by proteins is a very active field. It is expected that an increasing number of high-resolution transmembrane structures will be forthcoming in the next few years. These structures, together with kinetic measurements, are expected to give detailed information about the mechanism by which inorganic ions are transported across the cell membrane. The single-channel current-recording technique is ideally suited for studying channels that open because of the concentration gradient of inorganic ions and the resulting voltage changes. The change in transmembrane voltage determines whether or not a signal is propagated. Rapid chemical kinetic techniques with a 100-μsec time resolution, and suitable for investigations of ligand-gated ion channels on cell surfaces, are also now available. They are expected to provide additional information about ligand-gated ion channels and their mechanism of action. The ability to determine the effect of neurotransmitter concentration on the rate of transmembrane ion flux and, therefore, the change in transmembrane voltage is expected to provide important insight into how cells perceive, store, and transmit information. It is also expected to indicate how the receptor mechanism is changed by diseases of the nervous system and by the hundreds of drugs that affect the mechanism of these proteins. This information is expected to be essential in devising strategies for curing mental diseases and overcoming drug addiction.

SEE ALSO THE FOLLOWING ARTICLES

Bioenergetics • Bioinorganic Chemistry • Lipoprotein/Cholesterol Metabolism • Membrane Structure • Nucleic Acid Synthesis • Photochemistry, Molecular • Protein Structure • Protein Synthesis

BIBLIOGRAPHY

Doyle, D. A., Morais, C. J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1988). " The structure of the potassium channel: Molecular basis of K^+ conduction and selectivity," *Science* **280,** 669–677.

Hammes, G. G. (2000). "Thermodynamics and Kinetics for the Biological Sciences," Wiley Interscience, New York.

Hess, G. P., and Grewer, C. (1998). Development and application of caged ligands for neurotransmitter receptors in transient kinetic and neuronal circuit mapping studies. *In* "Methods in Enzymology" (G. Marriott, ed.), Vol. 291, pp. 443–474, Academic Press, New York.

Koester, J., and Siegelbaum, S. A. (2000). Ion channels. *In* "Principles of Neural Science" (E. R. Kandel, J. H. Schwartz, and T. M. Jessel, eds.) 4th edition, pp. 105–125, McGraw-Hill, New York.

Nelson, D., and Cox, M. M. (2000). Biological membranes and transport. *In* "Lehninger Principles of Biochemistry," 3rd edition, pp. 389–436, Worth Publishers, New York.

Palczewski, K., Kumasada, T., Hori, T., Behnke, C. A., Motoshima, J., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000). "Crystal structure of rhodopsin: A G protein-coupled receptor," *Science* **289,** 739–745.

Penner, R. (1995). A practical guide to patch clamping. *In* "Single-channel Recording" (B. Sakmann, and E. Neher, eds.), 2nd edition, pp. 3–31, Plenum Press, New York.

Subramaniam, S., and Henderson, R. (2000). "Crystallographic analysis of protein conformational changes in the bacteriorhopsin photocycle," *Biochim. Biophys. Acta* **11460,** 157–165.

Unwin, N. (2000). "Nicotinic acetylcholine receptor and the structural basis of fast synaptic transmission," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **355,** 1813–1829.

Lipoprotein/Cholesterol Metabolism

Alan D. Attie
University of Wisconsin

I. Lipid Absorption
II. Plasma Lipoprotein Structure
III. Chylomicron Metabolism
IV. VLDL Metabolism (Endogenous Triglyceride Metabolism)
V. IDL and LDL Metabolism
VI. Apolipoproteins Mediate Lipoprotein Metabolism
VII. The LDL Receptor
VIII. Familial Hypercholesterolemia
IX. How Do Sterols Regulate Gene Expression?
X. Other Lipoprotein Receptors
XI. HDL and "Reverse Cholesterol Transport"
XII. Tangier Disease and Familial Hypoalphalipoproteinemia
XIII. Lp(a) and apo(a)
XIV. CommonIsoforms of Apolipoprote in E(apoE)
XV. Apo-E and Alzheimer's Disease
XVI. Familial Combined Hyperlipidemia and Hypertriglyceridemia
XVII. Treatment of Lipoprotein Disorders
XVIII. Final Perspective

GLOSSARY

ABCA1 A membrane lipid transport protein involved in cholesterol and/or phospholipid efflux from cells. Its action is necessary for the extracellular assembly of lipoprotein particles.

Apo-A1 An apolipoprotein principally associated with HDL, an activator of lecithin cholesterol:acyltransferase. It interacts with cells to mediate delivery of cholesterol ester from HDL particles.

Apo-B100 An apolipoprotein associated with VLDL and LDL particles, synthesized in the liver; it is a ligand for the LDL receptor.

Apo-B48 An apolipoprotein associated with chylomicrons, synthesized in the intestine; it is a truncated form of apo-B100 and does not bind to the LDL receptor.

Apo-E An apolipoprotein principally associated with VLDL and chylomicrons; responsible for the receptor-mediated clearance of IDL and chylomicron remnants. It is a ligand for most members of the LDL receptor superfamily. The apo-E4 isoform is associated with increased risk of Azheimer's disease.

Apolipoproteins The protein components of plasma lipoproteins.

Bile acids Detergent-like molecules formed from cholesterol. They are secreted by the liver and, together withcholesterol and phospholipids, form bile. Bile forms micelles which emulsify lipids in the intestinal lumen aiding in their absorption.

Cholesterol ester transfer protein (CETP) A blood-borne protein that catalyzes the exchange of triglycerides in VLDL for cholesterol esters in HDL.

Chylomicron remnants Chylomicron particles that have been depleted of triglyceride after the lipoprotein lipase-mediated hydrolysis of their triglycerides.

Chylomicrons Lipoprotein particles produced in the intestine to package and secrete dietary lipids. Chylomicrons are secreted into the mesenteric lymph.

Familial hypercholesterolemia An elevation in LDL cholesterol due to mutations at the LDL receptor locus.

Familial hypoalphalipoproteinemia A deficiency in HDL due to mutations at the ABCA1 locus.

Fibric acid A class of drugs used to decrease the rate of VLDL triglyceride secretion.

Gallstones Large crystals, usually made of cholesterol, that form in the biliary tract and/or gall bladder when cholesterol levels in bile are too high relative to phospholipid and bile acids.

HDL High density lipoprotein; carries about 20% of plasma cholesterol. Its levels negatively correlate with risk of coronary heart disease. Is thought to mediate "reverse cholesterol transport."

HMG-CoA reductase A major rate-limiting step in the cholesterol biosynthetic pathway; catalyzes the conversion of HMG-CoA to mevalonate.

IDL Intermediate density lipoprotein, also called "VLDL remnant"; the VLDL particle that has been depleted of triglyceride through the action of lipoprotein lipase.

LDL Low density lipoprotein; carries approximately two-thirds of plasma cholesterol. Its levels are positively correlated with risk of coronary heart disease. It is formed in the circulation from the catabolism of VLDL.

LDL receptor Receptor expressed in most tissues and mainly necessary for normal clearance of LDL from the bloodstream.

Lipoprotein lipase Enzyme that catalyzes the hydrolysis of VLDL and LDL triglycerides to free fatty acids and glycerol; present on the luminal surface of the capillary endothelium.

Lipoproteins Particles that transport lipids in the bloodstream.

Lp(a) An LDL-like lipoprotein particle in which apo-B100 is connected through a disulfide linkage to apo(a), a plasminogen-related protein.

Microsomal triglyceride transfer protein (MTP) A protein in the endoplasmic reticulum necessary for secretion of VLDL and chylomicrons.

Nicotinic acid A coenzyme precursor that at pharmacological doses helps to lower triglyceride levels and raise HDL levels.

Reverse cholesterol transport A hypothetical pathway by which cholesterol is transported from extrahepatic-tissues to the liver via HDL.

Scavenger receptors Receptors that bind to a wide range of molecules, including modfied forms of LDL. They are involved in the accumulation of cholesterol in macrophages and smooth muscle cells in the arterial wall.

SREBP cleavage activating protein (SCAP) A cholesterol-sensing protein that, in the absence of cholesterol, escorts SREBP from the endoplasmic reticulum to the Golgi, where it is proteolytically activated.

Statins Cholesterol-lowering drugs that act by inhibiting cholesterol synthesis and upregulating the LDL receptor.

Sterol responsive element binding protein (SREBP) A transcription factor that postively regulates sterol-responsive genes, is attached to the endoplasmic reticulum membrane, and is released through a cholesterol-sensitive proteolysis step.

Tangier disease A severe HDL deficiency syndrome caused by homozygous mutations at the ABCA1 locus.

VLDL Very low density lipoprotein; a triglyceride-rich lipoprotein assembled in the endoplasmic reticulum and Golgi of hepatocytes and then secreted into the bloodstream. While in the circulation, it gives rise to IDL and then LDL after it loses triglyceride.

ANIMALS TRANSPORT LIPIDS in an aqueous environment at concentrations up to one million times their solubility in water. They accomplish this task by surrounding water-insoluble lipids with amphipathic lipids and proteins to form plasma lipoproteins. A major part of lipid transport is to supply energy for muscle contraction and to deliver lipids to adipose tissue for storage. Disorders in lipoprotein metabolism are a major risk factor for premature coronary heart disease throughout the world. These disorders arise from dysfunction in apolipoproteins, particular enzymes, and lipid transfer proteins, and also secondary to disorders in carbohydrate metabolism.

I. LIPID ABSORPTION

Animals absorb and transport large quantities of lipids. A major challenge is posed by the solubility characteristics of lipids. Because they are essentially insoluble in water, they must be packaged in order to move through an aqueous environment. Since many lipids are amphipathic (have water-soluble and water insoluble moieties), they have detergent-like properties that can be harmful to membranes.

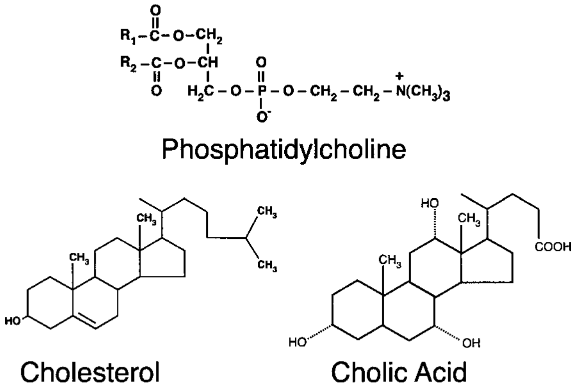


FIGURE 1 The major lipid components of bile. Bile acids are effective detergents and, together with phophatidylcholine and cholesterol, form micelles in the intestinal lumen. These micelles solubilize lipids and aid in their absorption by the intestinal mucosal cells.

Dietary cholesterol enters the intestinal lumen and is solubilized in a bile acid micelle. Cholesterol is otherwise quite insoluble in water (solubility limit $\approx 1\ \mu$g/liter).

The transport of lipids into the intestinal epithelial cells requires their solubilization in bile acid micelles. First, the solubilization facilitates the hydrolysis of fatty acid ester bonds by the intestinal lipases. Second, the transport into the enterocytes requires the formation of a properly structured bile acid micelle.

Bile is comprised of three lipid components: (1) bile acids, (2), cholesterol, and (3) phosphatidylcholine (PC; lecithin; Fig. 1). An excess of cholesterol relative to the two biliary amphipathic lipids (PC and bile acids) can lead to the formation of cholesterol precipitates, more commonly known as *gallstones*.

It is important to remember that ordinarily the major component of dietary fat is always triglyceride, not cholesterol. For example, milk and butter have very little cholesterol but are very high in triglyceride. Triglyceride (and other glycerolipids, such as phospholipids) are hydrolyzed in the intestinal lumen to yield monoglycerides and free fatty acids. These lipolysis products are then absorbed by the intestinal epithelial cells and resynthesized as triglycerides, phospholipids, and cholesterol esters. Thus, the intestine mediates both lipolysis (in the lumen) and re-esterification (within the epithelial cells) of dietary lipids.

The ability of the intestine to re-esterify monoglycerides and cholesterol is essential for net lipid absorption. This maintains a gradient that drives net lipid absorption from the intestinal lumen. Indeed, pharmaceutical companies have sought to develop inhibitors of cholesterol absorption that function by inhibiting the intestinal enzyme responsible for cholesterol esterification, acyl-CoA:cholesterol acyltransferase (ACAT).

Although the intestine has a large capacity for lipid absorption and esterification, it is not a lipid storage organ. It must therefore package and export absorbed lipids in order that they not accumulate. To accomplish this function, the intestine assembles the lipids into specialized particles called *plasma lipoproteins*.

II. PLASMA LIPOPROTEIN STRUCTURE

Plasma lipoproteins are uniquely endowed with the ability to transport large quantities of water-insoluble lipids through an aqueous environment. This because the nonpolar lipids (triglyceride and cholesterol ester; Fig. 2) are "buried" in the core of the lipoprotein, surrounded by a monolayer of amphipathic lipids, phospholipid, and unesterified cholesterol (Fig. 3).

In addition to core and surface lipids, lipoproteins carry proteins termed *apolipoproteins* (Tables III and IV). These proteins stabilize the lipoprotein particles and carry out particular functions such as receptor recognition and activation of particular enzymes.

Lipids are less dense than water. Consequently, because they are complexed with lipid, plasma lipoproteins tend to float when plasma is subjected to ultracentrifugation (Table I). In contrast, other blood proteins sediment in the centrifuge. Lipoproteins float at distinct buoyant densities and are named according to their flotation behavior. The lipoprotein classes are (Table II) *chylomicrons*, very-low density lipoprotein (*VLDL*), low density lipoprotein (*LDL*), and high density lipoprotein (*HDL*). Chylomicrons and VLDL are primarily triglyceride carriers, while LDL and HDL are primarily cholesterol (mostly cholesterol ester) carriers.

Key Points about the Distribution of Apolipoproteins

- Only HDL has no apo-B.
- For practical purposes LDL can be considered to have *only* apo-B.
- The other apolipoproteins overlap and, as described below, readily transfer between the various lipoprotein particles while circulating in the bloodstream.

It is useful to make a distinction between *exogenous* and *endogenous* lipid transport. Exogenous lipid transport refers to dietary fat. Endogenous lipids are those synthesized by the liver and adipose tissue from substrates that have already been absorbed and metabolized in these tissues. The exogenous pathway refers to the absorption of lipids in the intestine and the metabolism of *chylomicrons*. Lipids from chylomicrons can eventually mix with the endogenous lipid pools in the liver and adipose tissue.

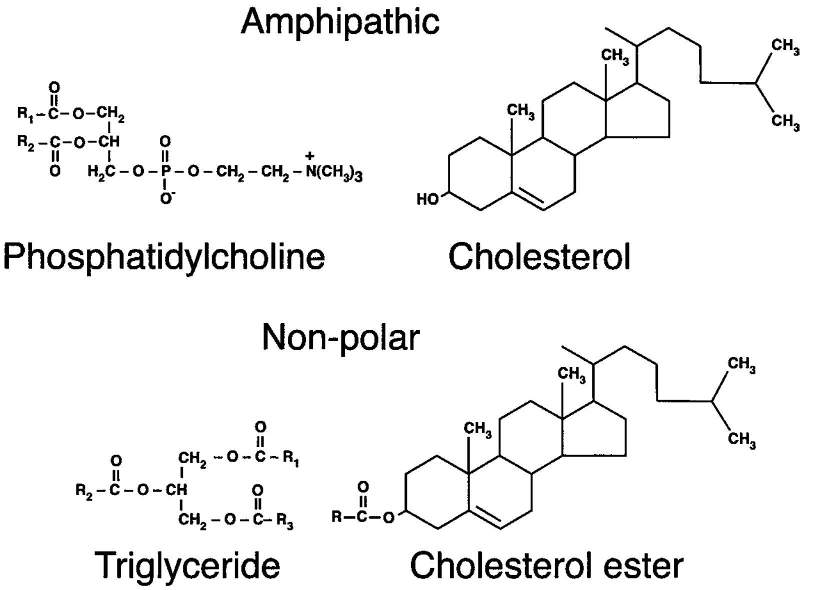


FIGURE 2 Lipids of plasma lipoproteins. Virtually all of the triglyceride and cholesterol ester of a lipoprotein is in the interior. Some unesterified cholesterol can also exist in the interior. However, all of the phospholipid is at the surface of the particle and surrounds the hydrophobic core.

III. CHYLOMICRON METABOLISM

Since the intestine is primarily an absorptive organ, it must have the means of exporting newly absorbed lipids. The enterocyte re-esterifies fatty acids and monoglycerides to form triglycerides and phospholipids. Absorbed cholesterol is esterified to form cholesterol esters. Under ordinary circumstances, the vast majority of the core lipids in the chylomicron are triglycerides; however, after a cholesterol-rich meal, the intestine manufactures cholesterol ester-rich chylomicron particles. Triglycerides and cholesterol esters are then packaged into the core of chylomicrons, which are secreted into the lymphatics. By secreting chylomicrons into the lymphatics, they gain entrance into the general circulation via the thoracic duct. This guarantees that extrahepatic tissues, principally adipose tissue and muscle, are the first to be exposed to the newly secreted chylomicrons—if chylomicrons were secreted directly into the bloodstream, they would first be delivered to the liver via the portal vein.

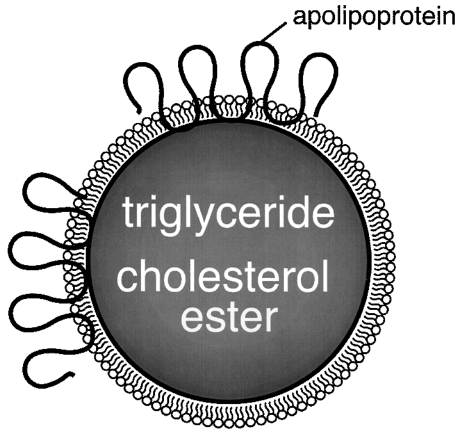


FIGURE 3 The domain structure of a plasma lipoprotein. The nonpolar lipids triglyceride and cholesterol ester are surrounded by the amphipathic lipids phospholipid and cholesterol. The latter are stabilized by apolipoproteins. These proteins have amphipathic α-helix and amphipathic β-sheet secondary structures.

Chylomicrons are very large (up to 1 μm in diameter). Thus a plasma sample containing chylomicrons is milky in appearance. A sample of fasting plasma is typically clear in appearance, even if there is an elevation in LDL particles—LDL particles are not large enough to scatter light.

In terms of *net* transport, the bulk of lipid flux is triglyceride $\rightarrow$ adipose tissue and muscle. How is the triglyceride

TABLE I Buoyant Density and Size of Plasma Lipoproteins

Class	Density (g/ml)	Diameter (nm)
Chylomicrons	0.93	75–1200
VLDL	0.93–1.006	30–80
IDL	1.006–1.019	25–35
LDL	1.019–1.063	18–25
HDL2	1.063–1.125	9–12
HDL3	1.125–1.210	5–9

TABLE II Chemical Composition of Plasma Lipoproteins[a]

Class	Chol	PL	Protein	TG	CE
Chylomicrons	2	7	2	86	3
VLDL	7	18	8	55	12
IDL	9	19	19	23	29
LDL	8	22	22	6	42
HDL2	5	33	40	5	17
HDL3	4	35	55	3	13

[a] Values represent percent of dry mass. Chol, cholesterol; PL, phospholipids; TG, triglycerides; CE, cholesterol esters.

transferred from the chylomicron particle to these tissues? As was the case with the transfer of lipids from the intestinal lumen to the intestinal epithelial cells, this transfer begins with a lipolytic reaction to convert an oily lipid, triglyceride, into an amphipathic lipid, free fatty acid. In this case, the reaction is catalyzed by *lipoprotein lipase*, an enzyme that resides at the luminal surface of the capillary endothelium in adipose tissue and muscle.

Main Features of Chylomicron Metabolism

- Chylomicrons are exclusively synthesized in the intestine.
- They are the package of dietary fat. Thus, few chylomicrons are made in fasted people.
- They are usually >80% triglyceride.
- They are secreted into the *lymphatics*, rather than the bloodstream.
- They are acted upon by *lipoprotein lipase* on the surface of adipose and muscle capillaries.
- After depletion of their triglyceride via the lipoprotein lipase reaction, the resulting remnants are cleared from the circulation.
- Chylomicrons are *not* converted into LDL.

Lipoprotein lipase catalyzes the complete hydrolysis of triglycerides to free fatty acid and glycerol. This reaction occurs in the bloodstream, while the chylomicron particle is in proximity to the endothelial surface of the capillary wall. Cholesterol ester is *not* a substrate for lipoprotein lipase. Therefore, with the selective loss of triglycerides, the particle becomes more enriched in cholesterol ester. The free fatty acids are then taken up and oxidized in muscle or re-esterified and stored as triglyceride droplets in adipose tissue. The abundance of the adipose form of lipoprotein lipase is increased by insulin, thus promoting uptake of free fatty acids and storage as cytoplasmic triglyceride droplets (see Fig. 4).

A chylomicron that has been depleted of its triglyceride core is termed a *chylomicron remnant*. Chylomicron

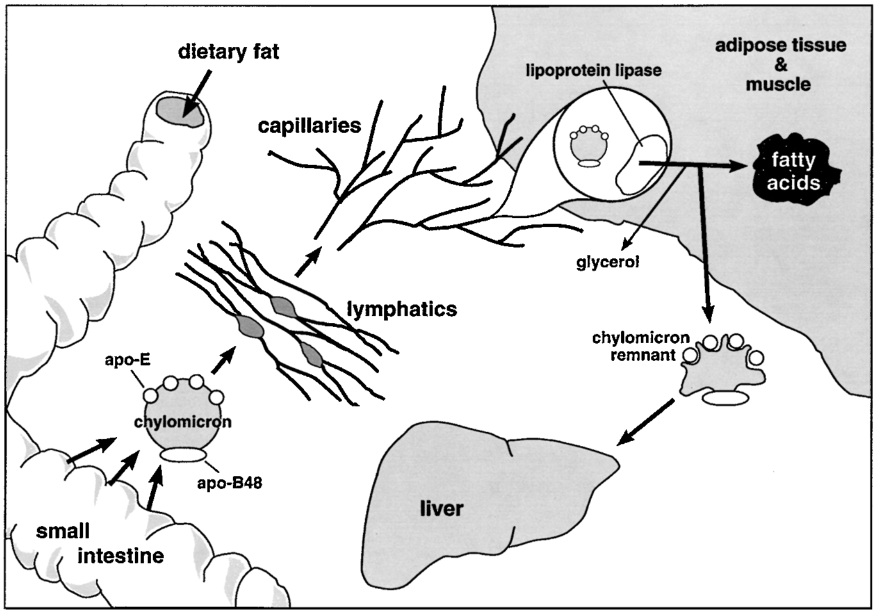


FIGURE 4 Chylomicron pathway. The intestine secretes chylomicron particles into the lymphatics. They gain entrance into the general circulation through the thoracic duct. Lipoprotein lipase, on the luminal surface of adipose and muscle capillary endothelial cells, hydrolyzes the triglyceride core to free fatty acids and glycerol. The free fatty acids are re-esterified and stored as triglycerides in adipose tissue or undergo β-oxidation in muscle. The lipid-depleted chylomicrons, chylomicron remnants, are cleared by the liver through a pathway that depends on apolipoprotein-E as a ligand for cellular receptors.

remnants are cleared from the circulation, primarily by the liver.

IV. VLDL METABOLISM (ENDOGENOUS TRIGLYCERIDE METABOLISM)

VLDL assembly and secretion is similar to the corresponding pathway for chylomicrons. Triglycerides and cholesterol esters are packaged into the core of the lipoprotein particle. However, in contrast to intestinal chylomicron secretion, hepatocytes secrete VLDL *directly* into the bloodstream. In the bloodstream, VLDL is acted upon by lipoprotein lipase, delivering its triglyceride cargo to muscle and adipose tissue. The resulting VLDL remnant particle, also termed *IDL* (intermediate density lipoprotein), is further metabolized as discussed below.

In both the liver and the intestine, triglyceride incorporation into lipoprotein particles requires the action of microsomal triglyceride transfer protein (*MTP*). MTP resides in the lumen of the endoplasmic reticulum (ER) and facilitates the transfer of triglycerides and cholesterol ester from the cytoplasmic side of the ER to the interior of the lipoprotein particle. Mutations in MTP cause abetalipoproteinemia, an inability to secrete chylomicrons and VLDL.

In humans, the liver is a major lipogenic tissue. The liver is able to transform excess carbohydrate or protein into fat (remember, when we eat too much of anything we get fat!). Fatty acid substrates for hepatic triglyceride production are derived from three sources (Fig. 5): (1) a continuous supply of albumin-bound fatty acid to the liver, primarily from adipose tissue triglyceride stores (after a meal this source drops, due to the antilipolytic action of insulin), (2) dietary fat already transported in chylomicrons delivered to the liver, and (3) carbohydrate in excess of the liver's capacity for storage as glycogen.

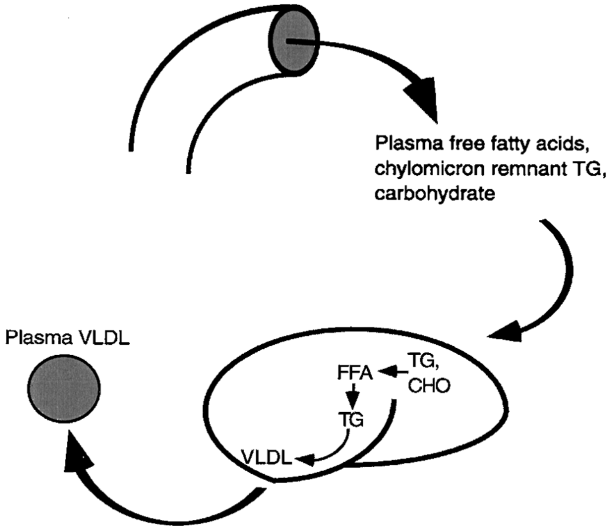


FIGURE 5 Sources of fatty acid for hepatic triglyceride synthesis. (1) Adipose tissue lipolysis, catalyzed by hormone-sensitive lipase, provides fatty acids that travel through the bloodstream to the liver. (2) Chylomicron remnants still carry some triglyceride and are cleared from the circulation by the liver. (3) Carbohydrate is converted to fatty acids when glycogen stores are maintained.

V. IDL AND LDL METABOLISM

Unlike chylomicron remnants, IDL has two competing metabolic fates: (1) uptake by the liver and (2) further processing to become LDL (Fig. 6).

Since IDL can be cleared by the liver or can be processed to become LDL, this branch point represents an important stage where LDL concentrations can be regulated. Inefficient clearance of IDL tends to lead to increased LDL production.

LDL is formed in the bloodstream through the catabolism of VLDL at the surface of blood vessels. In addition, VLDL is a triglyceride-rich lipoprotein, while LDL is cholesterol rich. The production of a cholesterol-rich lipoprotein from a triglyceride-rich lipoprotein occurs by selective removal of triglyceride from VLDL.

In summary, dietary fat is packaged into chylomicrons *in the intestine*. Dietary cholesterol is also packaged into chylomicrons, but a substantial fraction is delivered to the liver via chylomicron remnants. This cholesterol can then compete for the core of VLDL and appear in the bloodstream as cholesterol ester-enriched VLDL particles. Excess substrate in any form is converted into TG for export by the liver. For example, in some people, diets high in simple carbohydrates (e.g., fructose and sucrose) can lead to hypertriglyceridemia.

VI. APOLIPOPROTEINS MEDIATE LIPOPROTEIN METABOLISM

Chylomicrons and VLDL carry many apolipoproteins, among them apo-B and apo-E. Apo-B is a large (MW = 516 kDa) protein stably associated with the lipoprotein particle. The intestine produces a truncated form of the apoB, termed *apo-B48*, which is missing the carboxy-terminal half of apoB. The mechanism for the truncation involves a unique RNA editing event in the intestine that changes a Gln codon (CAA) to a STOP codon (UAA). Since only VLDL is converted to LDL, this means all apo-B in LDL is apo-B100 (Fig. 7).

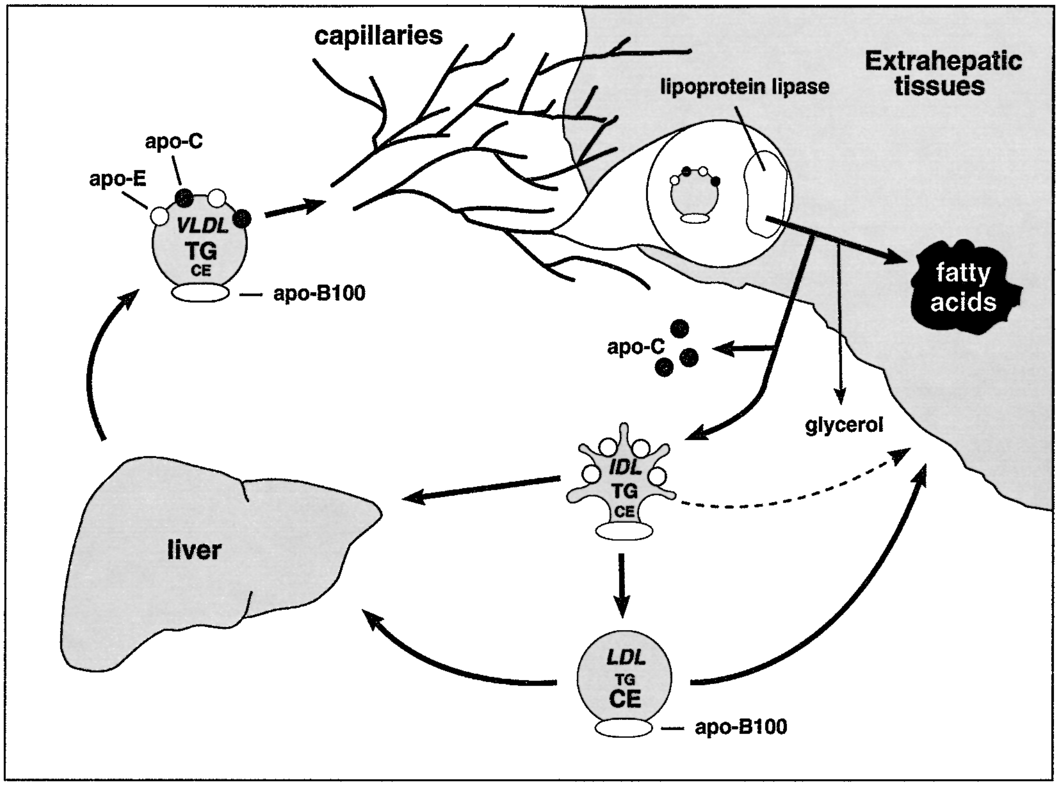


FIGURE 6 The VLDL → IDL → LDL pathway. VLDL is secreted by the liver directly into the bloodstream. Like chylomicrons, VLDL triglyceride is a substrate for lipoprotein lipase and is hydrolyzed while at the luminal surface of adipose tissue and muscle capillaries. The VLDL remnant (IDL), unlike the chylomicron remnant, can give rise to LDL. Some IDL is also directly cleared by the liver. LDL is cleared by the liver and by extrahepatic tissues.

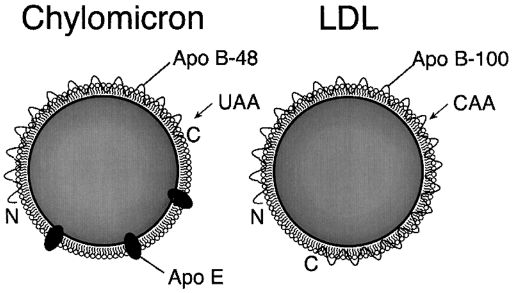


FIGURE 7 Chylomicrons contain a different form of apoB than VLDL or LDL. In the intestine, an RNA editing event introduces a stop codon in apo-B, resulting in a truncated protein product, apo-B48. VLDL is secreted with full-length apoB, apo-B100, and thus gives rise to LDL particles with apo-B100. The receptor-binding domain of apoB is at the C-terminal half of the protein, thus apo-B48 cannot bind to the LDL receptor; chylomicrons depend upon apo-E for receptor binding. Note: the particles are not drawn to scale; chylomicrons are about five times larger than LDL particles.

Main Features of VLDL and IDL Metabolism

- The liver is an important lipogenic tissue and exports newly synthesized triglyceride in VLDL particles.
- VLDL is synthesized in the liver and secreted *directly* into the bloodstream.
- As occurs with chylomicrons, VLDL is acted upon by lipoprotein lipase, on the surface of adipose and muscle capillaries. Unlike chylomicrons, VLDL remnants (IDL) are both cleared from the circulation *and* converted to LDL. This branch point is a factor that determines the rate of LDL production.

Whether or not a particle carries apo-B48 or apo-B100 has physiological importance. The carboxy-terminal half of apo-B is required for receptor recognition. Thus, apo-B100 can bind to cellular receptors (described below); apo-B48 cannot. For this reason, chylomicron remnants cannot be cleared from the circulation via apo-B48. Instead, chylomicron remnant clearance is mediated by another receptor ligand, apo-E. Indeed, genetic deficiency of apo-E leads to massive accumulation of cholesterol ester-rich chylomicron remnants and IDL in the bloodstream.

If chylomicrons and VLDL contain apo-E, why are they not cleared before they are acted upon by lipoprotein lipase? One explanation is that the apo-E on chylomicrons and VLDL is masked by another protein, apoC-III. During

lipolysis, the apoC-III protein falls off the particle, thereby exposing apo-E and permitting it to mediate particle clearance.

Some Key Points

- In terms of mass movement, export of triglyceride (TG) from the liver and intestine is where the action is. This means *net* movement to extrahepatic cells.
- Most of the plasma TG (in the fasting state, when no chylomicrons are present) is carried in VLDL. Thus, if fasting plasma TG is high, VLDL is high.
- Dietary fat is packaged into chylomicrons in the intestine. Excess endogenous substrate is converted into TG for export by the liver.
- Chylomicrons and VLDL are depleted of TG in the circulation, thus accomplishing their mission of delivering TG to peripheral cells.
- Chylomicron remnants are *not* converted to LDL; they are cleared by the liver.
- Only VLDL can be converted to LDL.
- In humans, most of the plasma cholesterol is carried in LDL. If plasma cholesterol is elevated without elevation of plasma TG, then LDL is high.

VII. THE LDL RECEPTOR

The discovery of the LDL receptor pathway by Michael S. Brown and Joseph L. Goldstein represents the most significant triumph in the field of atherosclerosis research. In an extraordinary collaboration begun in 1972, they discovered that cells possess a high-affinity receptor that binds to the apo-B100 moiety of LDL. (They were awarded the Nobel Prize in 1985.)

Binding of LDL to its receptor results in rapid endocyhtosis and the formation of an endocytic vesicle (Fig. 8). The LDL and the receptor separate while in this vesicle and part ways; the receptor recycles and returns to the cell surface, while the LDL particle is delivered to the lysosome, where the protein and lipid moieties are degraded.

Hydrolysis of LDL cholesterol esters in the lysosome results in the release of free cholesterol, which exits the lysosome and exerts three important regulatory functions:

1. It suppresses cellular cholesterol synthesis by reducing the levels of the rate-limiting enzymes in the cholesterol biosynthetic pathway, principally 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMG-CoA reductase).
2. It enhances the re-esterification of cholesterol for storage in a cytoplasmic lipid droplet.

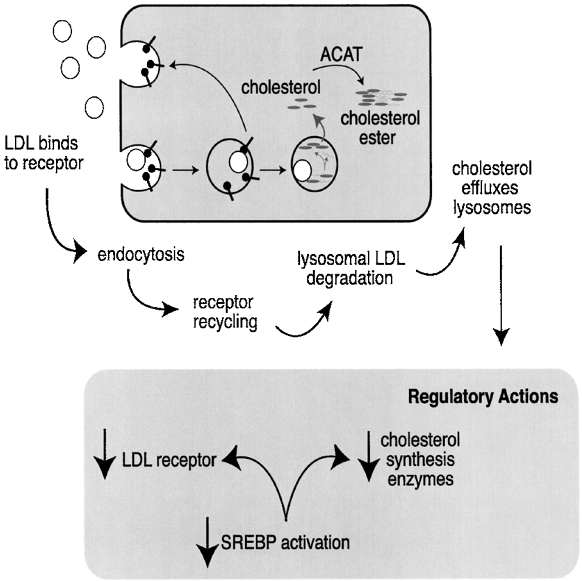


FIGURE 8 The LDL receptor pathway. LDL is internalized via receptor-mediated endocytosis. The endosomes are a sorting compartment; the receptor recycles to the plasma membrane, while the LDL is delivered to the lysosomes, where the cholesterol esters are hydrolyzed by lysosomal lipases. The free cholesterol then exits the lysosome and is able to inhibit *de novo* cholesterol synthesis by reducing the abundance of several cholesterol biosynthetic enzymes (e.g., HMG-CoA reductase) and the LDL receptor. Cells protect themselves from cholesterol toxicity by re-esterifying cholesterol to form a cytoplasmic cholesterol ester droplet. [From Brown, M. S., and Goldstein, J. L. (1986). *Science* **232,** 34–47.]

3. It inhibits production of new LDL receptor, thus diminishing the further supply of cholesterol to the cell.

The LDL receptor pathway assures a constant steady-state level of cellular cholesterol. This is accomplished both by adjusting cellular cholesterol synthesis according to ambient LDL levels and by altering LDL receptor number to limit the amount of LDL getting into cells. Like free fatty acids, unesterified cholesterol can be toxic to cells. The formation of cholesterol esters protects cells from cholesterol toxicity.

About two-thirds of LDL is catabolized by the liver. The rest is cleared by just about all other tissues. Steroid-producing tissues are especially active in LDL uptake. Adrenal cells (and presumably ovarian and testicular cells) do not synthesize cholesterol at rates sufficient to support high rates of steroidogenesis. They supplement their cholesterol supply by consuming cholesterol carried on LDL and HDL.

The level of LDL receptor activity is affected by the steady-state level of cholesterol in a cell. Thus, any factors

that increase or decrease the cholesterol level of a cell will affect the rate of LDL clearance from the circulation. This means that nutritional factors (proportion and type of dietary fat), hormonal status, pharmacological factors (drugs that inhibit cholesterol synthesis), and agents that affect bile acid metabolism all affect plasma cholesterol by influencing the level of expression of the LDL receptor.

VIII. FAMILIAL HYPERCHOLESTEROLEMIA

Brown and Goldstein studied LDL metabolism in cells from patients with a common metabolic inherited disorder called familial hypercholesterolemia (*FH*). People homozygous for this mutation have a 6- to 10-fold elevation of LDL levels, are born with detectable atherosclerosis, and usually do not survive childhood without a myocardial infarction. Heterozygotes have two- to fourfold elevations in LDL and suffer from coronary heart disease (CHD) during middle age (85% of FH heterozygotes have a heart attack before the age of 60.) FH is the most common inherited metabolic disorder in humans, with a gene frequency of 1 in 500, i.e., 1 in 500 people is a heterozygote for FH.

Brown and Goldstein discovered that FH cells from homozygote donors showed little or no LDL-binding activity. FH cells from heterozygotes possessed about 50% of normal activity. They concluded that mutations in the gene encoding the LDL receptor are the molecular basis for the FH disease. The inability to clear LDL through the normal LDL receptor pathway causes hypercholesterolemia atherosclerosis.

The loss of LDL receptor activity readily explains the inefficient clearance of LDL and the hypercholesterolemia of FH patients (Fig. 9). In addition to defective catabolism of LDL, there is also LDL overproduction for the following reason. IDL is also cleared through the LDL receptor. Diminished LDL receptor activity leads to prolonged circulation of IDL, giving it a greater opportunity to be converted to LDL. IDL is at an important branch point in lipoprotein metabolism; it can be directly cleared from the circulation or it can be further processed to become LDL. The LDL receptor can also regulate the secretion of VLDL. It promotes reuptake of newly secreted VLDL. Also, within the secretory pathway, the LDL receptor promotes the degradation of newly synthesized apo-B100.

The principal regulatory site in the cholesterol biosynthetic pathway is the step catalyzed by the enzyme 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase; Fig. 10). Regulation of HMG-CoA reductase by LDL cholesterol occurs principally through the regulation of the level of its mRNA. In analogy to genes regulated by steroid hormones, nuclear transcription factors bind to DNA sequences upstream from the reductase gene and regulate transcription. In addition to transcriptional regulation, cholesterol and certain other sterols diminish HMG-CoA reductase by a second mechanism; they hasten the degradation of the enzyme. The protein is located at the

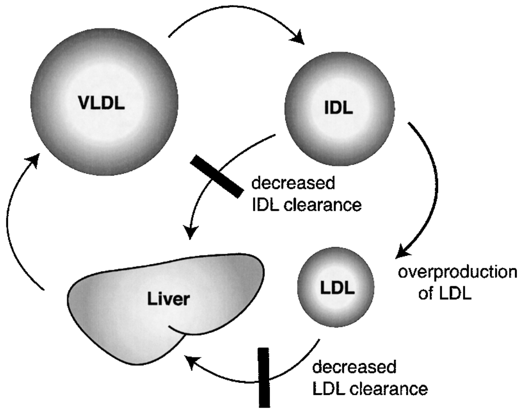


FIGURE 9 Kinetic mechanism for LDL overproduction in familial hypercholesterolemia. VLDL catabolism gives rise to IDL. IDL has two competing fates. It can be cleared by the liver or continue to be processed to become LDL. In the absence of the LDL receptor, IDL clearance is sluggish, thus a large proportion of IDL is converted to LDL.

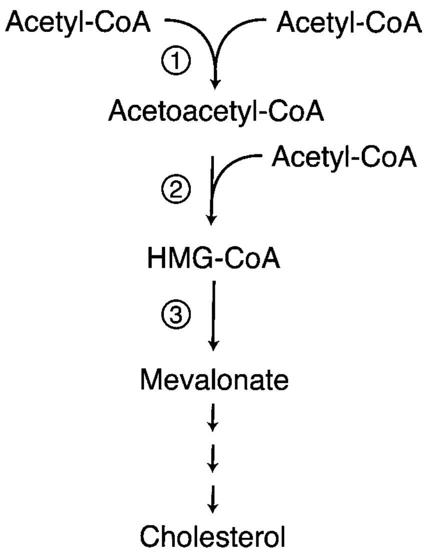


FIGURE 10 Regulated steps in cholesterol synthesis pathway. Step 1 is catalyzed by cytosolic acetoacetyl-CoA synthase. Steps 2 and 3 are catalyzed by HMG-CoA synthase and HMG-CoA reductase, respectively. The later two enzymes are transcriptionally regulated by SREBP. Cholesterol feeds back on its own synthesis by decreasing the abundance of enzymes 2 and 3. HMG-CoA reductase is the target of widely used cholesterol-lowering drugs known as "statins." Between mevalonate and cholesterol are more than 30 steps and branch points to nonsteroidal isoprenoid molecules.

ER membrane and has several transmembrane domains that are necessary in order for cholesterol to stimulate the degradation of the protein.

IX. HOW DO STEROLS REGULATE GENE EXPRESSION?

Cholesterol regulates its own formation by inhibiting the transcription of several genes in the cholesterol pathway, most notably HMG-CoA synthase and HMG-CoA reductase. For many years it was also known that polyunsaturated fats decrease the level of cholesterol synthesis. Now we know how these regulatory events occur.

The transcription of the cholesterol-regulated genes is regulated by a regulatory region that is upstream (before the transcriptional start site) of these genes. A special DNA sequence termed *sterol responsive element (SRE)* determines the responsiveness of these genes to regulation by cholesterol. How does cholesterol inhibit the transcription of genes with SREs?

A transcription factor that binds to SREs is termed *sterol regulatory element binding protein (SREBP)*. This protein turns *on* the transcription of genes with SREs in front of them, thus is a positive transcription factor.

SREBP is found in the nuclear and endoplasmic reticulum membrane in an inactive form. To be activated, it must be cleaved from the membrane and released so that it can enter the nucleus and turn on transcription. The key to cholesterol regulation is that cholesterol (or more likely, a metabolite of cholesterol) inhibits this cleavage event. The "sensing" of cholesterol is carried out by another protein, *sterol cleavage activated protein (SCAP)*, a protein that binds to SREBP. In sterol-depleted cells, SCAP escorts SREBP from the ER to the Golgi, where it is activated by proteolytic cleavage. This transport step is blocked by sterols. Interestingly, unsaturated fatty acids also inhibit SREBP activation, thus explaining how they inhibit cholesterol synthesis (Fig. 11).

The LDL receptor is also regulated by an SRE. This explains why cholesterol downregulates the activity of the LDL receptor. Many genes in fatty acid and triglyceride synthesis are regulated by SREs. The list is growing; thus the importance of SREBP in physiology will be enlarged in the future.

The expression of SREBP is enhanced by insulin. This helps to understand how insulin promotes lipogenesis through through global activation of expression of numerous lipogenic enzymes. In some individuals on high carbohydrate diets, plasma VLDL levels rise, a consequence of an abnormally high rate of *de novo* lipogenesis. The hyperinsulinemia that accompanies insulin resistance (see Section XVI) is also associated with increased levels of VLDL. This might be a consequence of insulin-mediated stimulation of SREBP expression.

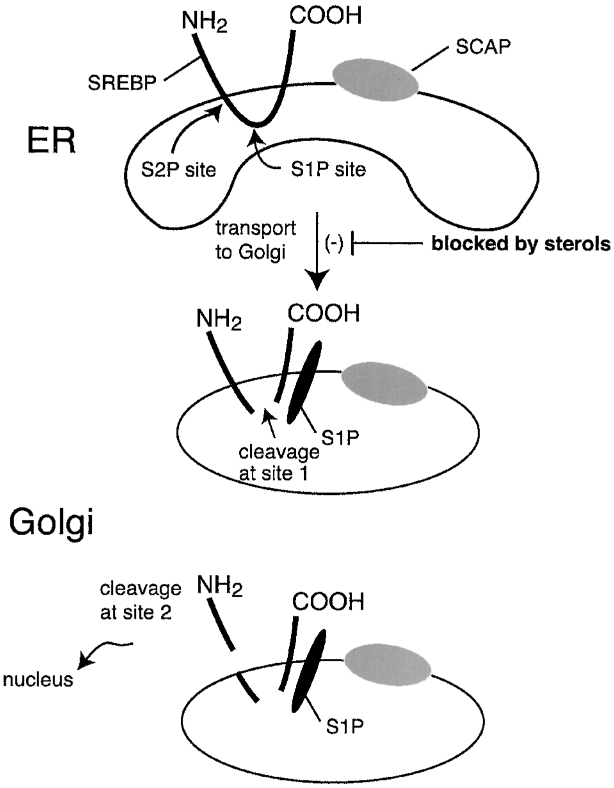


FIGURE 11 Transcriptional regulation of sterol-responsive genes. A transcription factor termed sterol regulatory element binding protein (SREBP) binds to the SREs and enhances transcription. However, the SREBP is held captive bound to the endoplasmic reticulum membrane. Only when it is released by proteolytic cleavage does it travel to the nucleus, where it regulates sterol-responsive genes. The protein traverses the membrane twice and is cleaved by the successive actions of two proteases. The proteolysis step occurs in the Golgi. Transport of SREBP to the Golgi requires a second protein, SCAP. The transport step is inhibited by cholesterol through a sterol-sensing function of SCAP. Thus, cholesterol regulates gene expression by controlling the activation of a membrane-bound transcription factor, SREBP.

Patients lacking the LDL receptor do not accumulate chylomicron remnants in the bloodstream. Since chylomicron remnant clearance is mediated by apo-E, it has been postulated that a separate receptor is responsible for chylomicron remnant clearance, a receptor that, in contrast to the LDL receptor, binds to apo-E, but not to apo-B100. Several additional members of the LDL receptor family have been identified (Table V). The first of these, the *LRP*, participates in chylomicron remnant clearance and plays a major role in that process when the LDL receptor is absent or dysfunctional.

TABLE III Properties of the Apolipoproteins

Apolipoprotein	Molecular weight (Da)	Concentration (mg/dl)	Tissue origin	Function
Apo-A1	28,300	90–130	Intestine, liver	LCAT activator
Apo-A2	17,000	30–50	Intestine, liver	Unknown
Apo-A4	46,000	10–30	Intestine, liver	Unknown
Apo-B100	516,000	80–100	Liver	Cholesterol, triglyceride transport, receptor recognition
Apo-B48	264,000	—	Intestine	Triglyceride, cholesterol transport
Apo-C1	6,500	4–7	Liver	LCAT activator
Apo-C2	8,800	3–8	Liver	LPL activator
Apo-C3	8,750	8–15	Liver	LPL inhibitor
Apo-D	33,000	10	Liver, intestine, pancreas, kidney, adrenals, brain	Unknown
Apo-E	35,000	3–6	Liver, macrophages, brain, Adrenal	Receptor recognition
CETP	74,000	—	Spleen, liver, small intestine, adrenal	Cholesterol ester transfer
Lp(a)	400–700	0.4–80	Liver, testes, brain	Unknown

X. OTHER LIPOPROTEIN RECEPTORS

A. LRP/α_2-Macroglobulin Receptor

The lipoprotein receptor-related protein (*LRP*)/α_2-macroglobulin receptor (LRP) has a mass about five times larger than the LDL receptor. It contains many of the same structural domains as the LDL receptor, including a cysteine-rich repeat domain that binds to lipoproteins. *In vitro* studies have shown that the LRP does not bind to apo-B100, but binds to VLDL particles that have been enriched in apo-E. The binding of LRP to its ligands is dependent upon cell surface *proteoglycans*. These are proteins containing sulfated carbohydrate residues.

In addition to binding lipoproteins, the LRP also binds to α_2-macroglobulin, a plasma protein that binds to proteases and, upon binding, is cleared by the liver. The LRP also binds to plasminogen activators and their inhibitors, such as tissue-type plasminogen activator, urokinase-type plasminogen activator, and their corresponding inhibitors.

TABLE IV Apolipoproteins of the Plasma Lipoproteins

Chylomicrons	A-1, A-2, A-4, B-48, C-1, C-2, C-3, E
VLDL	B-100, C-1, C-2, C-3, E, trace amounts of A-1, A-2, and A-4
LDL	90–98% apo B-100
HDL	A-1, A-2, sometimes, E

B. Cubulin

Cubulin is a 460-kDa protein localized to epithelial cells, such as those of the kidney. It plays a role in albumin reabsorption; its absence leads to loss of albumin in the urine. Cubulin also mediates the uptake of the major HDL protein apo-A1 by the kidney.

C. Megalin (gp330)

This receptor is a member of the LDL receptor superfamily. It is the antigen that elicits an autoimmune response leading to a disease called Heymann's glomerulonephritis. It interacts with cubulin and is thought to function together with cubulin in the uptake of apo-A1 by the kidney.

D. A VLDL Receptor (?)

A cDNA encoding another member of the LDL receptor family is expressed in muscle and adipose tissue, but *not* in liver. It binds to apo-E-containing lipoproteins *in vitro*, but not to LDL. Based on its tissue localization and binding properties, this protein has been postulated to be a receptor for triglyceride-rich lipoproteins.

E. CD-36

CD-36 is expressed in a variety of tissues, including adipocytes and macrophages. It functions in fatty acid transport and, like scavenger receptors, binds a variety

TABLE V Lipoprotein Receptors

Name	Ligands	Tissue locations
LDL receptor (LDLR)	LDL, apo-E, hepatitis C virus, Rous sarcoma virus subgroup A, human rhinovirus	Ubiquitously expressed
LDL receptor-related protein/ α_2-macroglobulin receptor (LRP)	Apo-E, lipoprotein lipase, hepatic lipase, thrombospondin, *Pseudomonas* exotoxin A, α_2-macroglobulin, receptor-associated protein (RAP), lactoferrin, t-PA, u-PA, t-PA:PAI-1, u-PA:PAI-1, elastase-a1-antitrypsin	Liver, brain, lung, adrenal, intestine, kidney, placenta, ovary, testis
VLDL receptor	apo-E, Reelin	Smooth muscle, brain, adipose, adipose, adrenal, testis, ovary, placenta, lung
gp330/megalin (VLDLR)	LDL, apo-J/clusterin, apolipoproteinH/β2-glycoprotein-I, PAI-1, prourokinase, lipopoprotein lipase, aprotinin, transcobalamin-vitamin B12, vitamin D-binding protein, retinol-binding protein, PTH, insulin, β2-microglobulin, epidermal growth factor, prolactin, lysozyme, cytochrome c, thyroglobulin, plasminogen, albumin, polybasic drugs, RAP, Ca^{2+}	Kidney, lung, thyroid, parathyroid, epididymis, ileum, placenta, thymus
Apo-E receptor R2 (apoER2; LR7/8B)	Apo-E, apo-B100, Reelin, RAP	Brain, testes, ovary, placenta

of ligands, including senescent neutrophils, collagen, and malaria-infected erythrocytes.

F. An HDL Receptor, SR-B1

Unlike LDL, HDL can deliver cholesterol to certain tissues (liver and steroidogenic tissues like adrenal, ovary, and leydig cells) without the HDL particle being internalized. The SR-B1 protein mediates the docking of HDL with cells and the delivery of cholesterol ester from the core of the HDL particle to the cells. This receptor is in the scavenger receptor family and is able to bind other lipoproteins in addition to HDL.

The discovery that LDL receptor deficiency in familial hypercholesterolemia is associated with premature atherosclerosis was initially paradoxical. A hallmark of atherosclerotic lesions is the formation of “foam cells”—macrophages and smooth muscle cells that accumulate cytoplasmic cholesterol ester droplets. If these cells lack the LDL receptor, how do they import excessive amounts of cholesterol?

Chemical modification of LDL (e.g., acetylation) abolishes its ability to bind to the LDL receptor. However, the modified LDL binds to other cell surface receptors, termed *scavenger receptors*. Unlike the LDL receptor, scavenger receptors are not downregulated by cholesterol. Like the LDL receptor, uptake through scavenger receptors leads to accumulation of LDL-derived cholesterol as cytoplasmic cholesterol ester droplets.

Are there any physiological counterparts to *in vitro* chemical modification of LDL? The polyunsaturated fatty acyl chains of LDL lipids can be oxidized, leading to cleavage and formation of chemically reactive short-chain fatty aldehydes. These aldehydes react with the protein moiety of LDL, apo B-100, converting it to a ligand for scavenger receptors. The discovery of LDL oxidation and its implications for atherosclerosis has ignited a strong interest in the potential role of dietary antioxidants (e.g., vitamins C and E) in the prevention of atherosclersosis.

Three scavenger receptors for oxidized LDL have been identified; SR-A1, SR-A2, and CD36 (Table VI). These receptors have also been implicated in the phagocytosis of damaged or apoptotic blood cells (e.g., lymphocytes and erythrocytes). CD36 has been implicated in fatty acid uptake into cells. The scavenger receptors can bind a wide array of molecules other than lipoproteins (e.g., polyanions, oligonucleotides, bacterial endotoxin, and crocidolite asbestos).

XI. HDL AND “REVERSE CHOLESTEROL TRANSPORT”

Epidemiological studies show a strong inverse relationship between HDL levels (HDL cholesterol or apo-A1) and risk of CHD. Even a small increase in HDL is significantly correlated with a reduction in the risk of premature heart disease.

Unlike VLDL and chylomicrons, HDL is formed from its protein and lipid components in the bloodstream and interstitial fluids. The major apolipoproteins of HDL are apo-A1 and apo-A2. These proteins are secreted from hepatocytes and intestinal epithelial cells independently and also as minor components of VLDL and chylomicrons. Apo-A1 and apo-A2 bind to phospholipids. Phospholipids are available from the surface of VLDL after lipolysis. In addition, cells are able to efflux phospholipids through the action of *ABCA1*.

ABCA1 is a membrane transport protein belonging to a large family of ATP-binding cassette proteins. These

TABLE VI Scavenger Receptor Family

Name	Ligands	Tissue locations
Class A		
SR-AI, SR-AII	Acetylated LDL, oxidized LDL, polyanions, crocidolite asbestos, bacterial endotoxin, lipoteichoic acid	Macrophages, (Kupffer cells, histiocytes, microglial cells), some endothelial cells (low level)
MARCO (SR-AIII)	Bacteria	Macrophages
Class B		
SR-BI	HDL, LDL, modified lipoproteins, anionic phospholipids, acetyl LDL	Highest expression in steroidogenic cells (adrenal, ovary, testis) and hepatocytes, lower level expression seen in absorptive epithelial cells of the proximal small intestine, lactating mammary gland, low levels observed in all cultured cells
CD36	HDL, LDL, modified lipoproteins, anionic phospholipids, fatty acids, collagen, malaria-infected erythrocytes	Adipose, macrophages, epithelial cells, monocytes, certain endothelial cells, platelets
Croquemort	Apoptotic cells	*Drosophila* macrophages
Class C		
SR-CI	Multiple polyanions, including modified LDLs, poly(I)	*Drosophila* macrophages
Class D		
Microsialin (CD68)	Modified lipoproteins	Macrophages, Kupffer cells, endothelial cells
Class E		
LOX-1 (SR-E1)	Oxidized LDL	Endothelial cells
Class F		
SREC	Oxidized LDL	Endothelial cells
FcgR2-B2	Oxidized LDL	Macrophages

proteins include the cystic fibrosis transmembrane receptor, the sulfonylurea receptor (a pancreatic β-cell protein involved in insulin secretion), and the multidrug resistance transporter proteins. It is thought that ABCA1 mediates the efflux of phospholipid and/or cholesterol from cells, thus making them available for association with apolipoproteins to form HDL. Its crucial role in this process was established by the discovery that two types of severe inherited HDL deficiency syndromes are caused by mutations in ABCA1 (see Fig. 12).

XII. TANGIER DISEASE AND FAMILIAL HYPOALPHALIPOPROTEINEMIA

Tangier disease is a rare recessive disorder in which patients have almost no HDL. Cholesterol ester accumulates in macrophages and macrophage-rich tissues like spleen and liver. Familial hypoalphalipoproteinemia (FHA) is a very common dominant disorder in which people have low HDL (typically <30 mg/dl) and suffer from premature heart disease even without an elevation in LDL.

Approximately 40% of patients with premature coronary heart disease have low HDL, making it the most common lipid disorder of heart disease patients.

Tangier disease and FHA are caused by mutations in the same gene, ABCA1. Tangier disease patients are homozygous (or inherit two different mutant alleles). FHA patients are heterozygous for mutations at the ABCA1 locus. ABCA1 mutations lower HDL because they prevent phospholipid and/or cholesterol from effluxing and becoming associated with apo-A1 and apo-A2. In the absence of sufficient lipid, apo-A1 is rapidly cleared by the kidneys.

Why do mutations in ABCA1 lead to premature heart disease? ABCA1 fulfills a rate-limiting step in the pathway by which cells get rid of cholesterol, and thus might be a critical protector against cholesterol overload. With the exception of hepatocytes, cells are unable to catabolize large quantities of cholesterol and must therefore protect themselves from cholesterol overload by expelling cholesterol to an appropriate extracellular carrier. It is interesting that ABCA1 is especially abundant in macrophages; macrophages can become engorged with cholesterol esters and form foam cells in the arterial wall. A mutation in ABCA1 might therefore predispose an individual to atherosclerosis by impeding cholesterol efflux.

Do HDL levels correlate with the rate of reverse cholesterol transport? Studies of the expression of the SR-B1 receptor provide an answer to this question. When HDL binds to SR-B1 at the plasma membrane of the hepatocyte, it unloads its cholesterol ester cargo. The lipid-depleted apo-A1 then is cleared from the circulation, in part by the kidney. High levels of SR-B1 therefore lead to low HDL

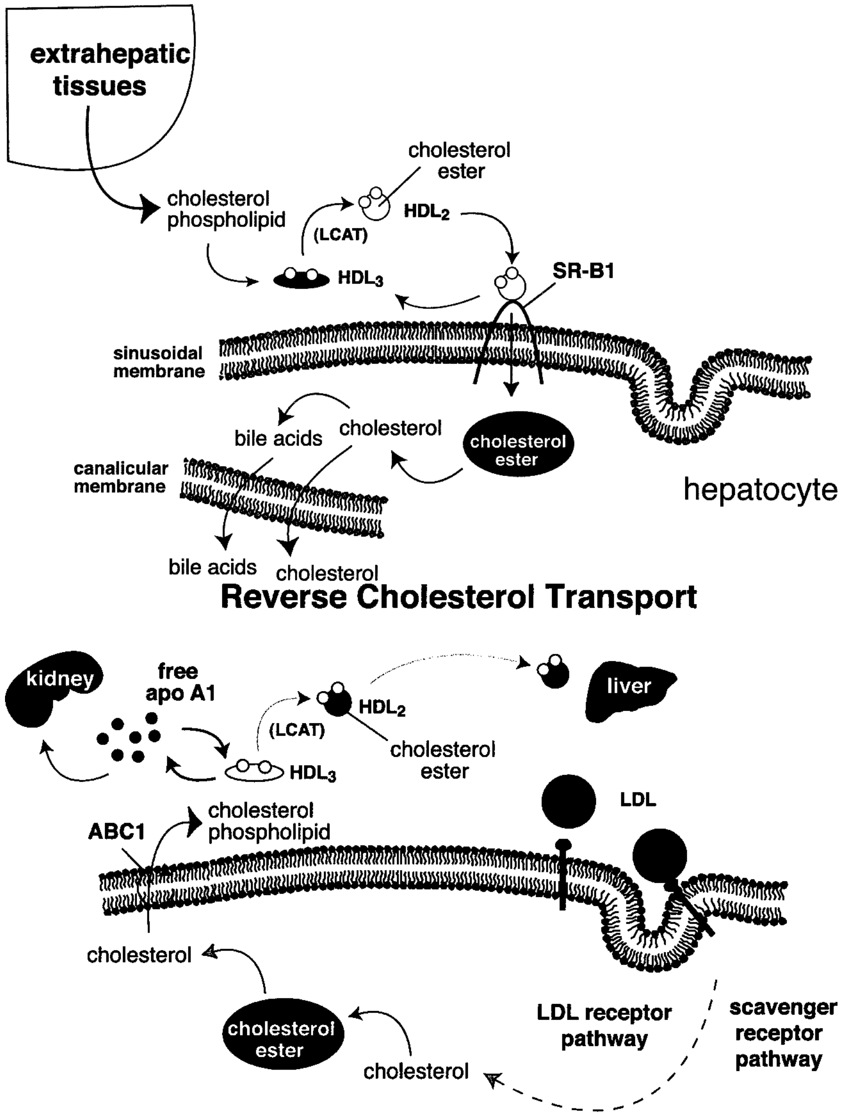


FIGURE 12 The reverse cholesterol transport hypothesis. Extraheptatic cells accumulate cholesterol through the uptake of LDL and modified forms of LDL through the LDL receptor, other members of the LDL receptor family, and/or scavenger receptors. These cells efflux cholesterol and phospholipids to the extracellular milieu through a process facilitated by a membrane transporter, ABC1. Free apo-A1 interacts with the phospholipid and cholesterol to form HDL3 particles. LCAT esterifies the cholesterol to form discoidal HDL2 particles which then interact with the SR-B1 receptor at the surface of hepatocytes. Through that interaction, cholesterol esters are selectively taken up into the hepatocytes and hydrolyzed. The free cholesterol is then secreted into the bile or converted to bile acids. Much of this cholesterol is then excreted in the feces.

levels. However, under these circumstances, the transport of cholesterol to the liver is enhanced. When the diet is shifted from saturated to polyunsaturated fat, HDL levels typically drop. This drop is attributable to a rise in SR-B1 expression in the liver. In short, a decrease in HDL can be caused by imparied reverse cholesterol transport (extrahepatic; Tangier disease) or by increased reverse cholesterol transport (liver; polyunsaturated fat).

HDL metabolism is intimately connected with triglyceride metabolism. Clinicians know this because patients with hypertriglyceridemia almost invariably have low HDL levels. The *cholesterol ester transfer protein (CETP)*-catalyzed transfer of cholesterol ester from HDL to VLDL occurs with reciprocal transfer of triglyceride to HDL (Fig. 13). Lipoprotein lipase and hepatic lipase (another cell-surface-bound lipase) catalyze the hydrolysis of HDL triglyceride, reducing the size of the HDL particles. Hepatic lipase can also hydrolyze HDL phospholipids. Smaller HDL particles (HDL_3) are removed from the circulation more quickly than larger HDL (HDL_2).

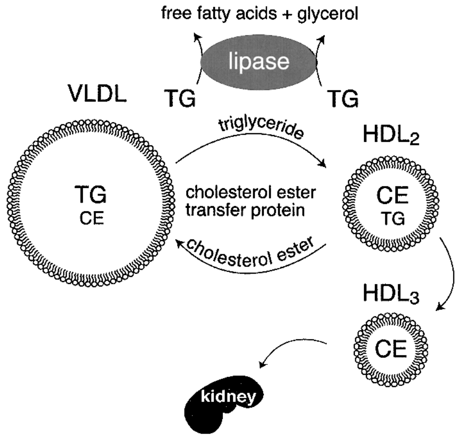


FIGURE 13 Inverse relationship between plasma triglyceride and HDL cholesterol levels. A higher level of VLDL correlates with lower HDL levels. Two processes simultaneously remove triglycerides from VLDL particles. First, lipoprotein lipase hydrolyzes the triglycerides to free fatty acids and glycerol. Second, cholesterol ester transfer protein (**CETP**) in the bloodstream catalyzes the exchange of triglyceride and cholesterol ester between VLDL and HDL, respectively. As HDL accumulates triglyceride, it is a substrate for lipoprotein lipase and hepatic lipase. This shrinks the HDL particles, causing them to be cleared by the kidneys.

The abundance of VLDL relative to HDL significantly influences HDL metabolism, probably due to enhanced exchange of triglyceride into the HDL particles. Individuals with genetically reduced levels of cholesterol ester transfer protein have extremely high HDL levels. Heavy exercise is also associated with increased HDL levels. Exercise increases the expression of muscle lipoprotein lipase. The resulting increase in VLDL lipolysis decreases the amount of triglyceride that can participate in the CETP lipid exchange process. This results in an elevation in HDL.

XIII. Lp(a) AND apo(a)

Apo(a) is a protein found covalently linked to apo-B100 in some LDL particles. Those LDL particles to which apo(a) is attached are called lipoprotein(a) or *Lp(a). Plasma levels of Lp(a) correlate with increased cardiovascular disease risk in most populations, but not in African-Americans.* The level of Lp(a) appears to be entirely genetically determined. The Lp(a) concentration is almost entirely related to the particular alleles of the Lp(a) gene expressed by an individual.

The Lp(a) gene is highly polymorphic (variable in structure). Thus, it appears that most individuals have different forms of the gene. The reason for this unusually high degree of genetic variability is that the gene comprises multiple repeat structures. Repeat structures lead to unequal crossing over during meiosis, causing the production of new variants of the gene in the offspring. The repeat structures are domains termed "kringle domains." The kringle domain that is repeated in the apo(a) gene is homologous to a kringle domain of plasminogen, an enzyme involved in the dissolution of fibrin clots.

Why is Lp(a) so strongly correlated with risk of CHD? Because of its homology with plasminogen, it has been suggested that Lp(a) competes with plasmin for binding to fibrin clots and therefore would tend to be antithrombolytic. Another hypothesis is that Lp(a) might be atherogenic because it has been shown *in vitro* to stimulate smooth muscle cell proliferation, a hallmark of the atherosclerotic process.

XIV. COMMON ISOFORMS OF APOLIPOPROTEIN E (apoE)

Apo-E occurs in three common isoforms, apo-E2, apo-E3, and apo-E4. They differ at amino acids 112 and 158 (Table VII). In apo-E4, both of these amino acids are arginine. In apo-E2, both amino acids are cysteine, and apo-E3 has Cys-112 and Arg-158. The presence of cysteine at amino acid 158 virtually abolishes the LDL receptor-binding activity of apo-E. Consequently, VLDL remnants with apo-E2 accumulate in the circulation. From 0.2–1.6% of individuals in different populations are E2/E2 homozygotes. A subgroup of E2/E2 individuals have an unusually severe form of hypercholesterolemia due to excessive remnant lipoproteins rather than high LDL. This disorder is called *Type III hyperlipidemia*.

Individuals with apo-E2 exhibit delayed clearance of chylomicron remnants. The delayed clearance of remnants means cholesterol delivery to the liver is reduced. This causes an upregulation of the LDL receptor, resulting in lower plasma LDL levels. Thus, the total cholesterol in E2/E2 individuals (except those with Type III disease) might be normal, even though they have a problem clearing chylomicron remnants.

Apo-E4 is associated with higher total cholesterol levels than apo-E2 or apo-E3. This has been attributed to the relatively high affinity of apo-E4 for VLDL particles. Enrichment of VLDL with apo-E results in enhanced clearance by the liver (through the LDL receptor) and greater downregulation of the LDL receptor, thus increased LDL levels.

TABLE VII Apo-E Isoforms

	Apo-E2	Apo-E3	Apo-E4
Amino acid 112	Cys	Cys	Arg
Amino acid 158	Cys	Arg	Arg
LDL receptor binding	<0.1%	Normal	Normal
LDL cholesterol	Low	Normal	High
VLDL cholesterol	High	Normal	Normal

TABLE VIII Major Lipoprotein Disorders

Disorder	Principal plasma abnormality	Clinical features[a]	Estimated frequency
Heterozygous familial hypercholesterolemia	↑LDL	Tendinous xanthomas, cornial arcus, premature CAD; family history of hypercholesterolemia	0.2% of general population; 5% of MI survivors under age 60
Familial combined hyperlipidemia	↑1/3 LDL only, ↑1/3 VLDL only, ↑1/3 LDL & VLDL, apoB overproduction common	Patients usually over age 30, often overweight, usually no xanthomas or premature CAD	0.5% of general population; 15% of MI survivors under age 60
Polygenic hypercholesterolemia	↑LDL	Premature CAD, no xanthomas	Unknown
Familial hypertriglyceridemia (200–500 mg/dl)	↑VLDL only	Patients often overweight, usually do not have xanthomas or premature CAD	1% of general population; 5% of MI survivors under age 60
Severe hypertriglyceridemia (>1000 mg/dl)	↑Chylomicrons & ↑VLDL	Patients usually middle age, obese, hyperuricemic, diabetic, at risk for pancreatitis	Unknown
Familial hypoalphalipoproteinemia	↓HDL	Premature CAD	∼40% of patients with premature CAD

[a] CAD, Coronary artery disease; MI, myocardial infarction.

XV. Apo-E AND ALZHEIMER'S DISEASE

Epidemiological studies have revealed a correlation between the apo-E4 allele and risk of developing Alzheimer's disease. About 80% of familial and 64% of sporadic Alzheimer's disease late-onset cases have at least one apo-E4 allele, compared to 31% of control subjects. The apparent risk is dose dependent. In one study, 91% of E4/E4 homozygous individuals from families with diagnosed Alzheimer's disease were affected, while only 48% of E3/E4 patients and approximately 20% of E2/E3 or E3/E3 patients were affected. Three hypotheses are being considered to explain this correlation. One suggests that apo-E4 associates more readily than the other apoE isoforms with the amyloid protein to form deposits. Another hypothesis states that apoE4 does not associate properly with a microtubule associated protein termed *tau*. A third hypothesis is that the different forms of apoE have distinct effects on the growth and extension of axons after nerve injury.

XVI. FAMILIAL COMBINED HYPERLIPIDEMIA AND HYPERTRIGLYCERIDEMIA

A very common disorder frequently associated with elevated VLDL and LDL or elevated VLDL only is termed familial combined hyperlipidemia (*FCHL*) (Table VIII). In order to make a diagnosis, the family of the patient must be screened. Some family members will display increases in VLDL, others in LDL, and some in both VLDL and LDL. The basic defect appears to involve overproduction of apo-B100. In some cases, this is accompanied by overproduction of triglyceride. However, the molecular basis for the defect and the responsible gene(s) have not been elucidated.

Hypertriglyceridemia is commonly associated with obesity and both Type I and Type II diabetes mellitus. (Note that fasting hypertriglyceridemia is synonymous with increased VLDL levels.) This disorder is almost invariably accompanied by low HDL levels (Fig. 11) and for this reason is a risk factor for premature heart disease. (Whether high VLDL by itself is a cardiovascular risk factor is controversial.)

In Type I diabetes mellitus, there is a severe deficiency (or total absence) of insulin due to an autoimmune attack on the cells that produce insulin, pancreatic β-cells. The absence of insulin produces a deficiency in adipose tissue lipoprotein lipase. This causes sluggish catabolism of VLDL and leads to hypertriglyceridemia. Another mechanism by which insulin deficiency promotes increased VLDL levels is the failure to inhibit the activity of adipose tissue hormone-sensitive lipase. This enzyme hydrolyzes cytoplasmic triglyceride droplets. The fatty acids then go to the liver, where they are re-esterified to form triglycerides. These triglycerides are exported on VLDL particles. Since adipose tissue-derived fatty acids are an important substrate for hepatic VLDL triglycerides, the failure to suppress adipose tissue lipolysis is an important contributor to the enhanced rate of VLDL triglyceride secretion.

Obesity is almost invariably associated with insulin resistance, a poor response of insulin's target tissues to an

appropriate physiological dose of insulin. To prevent diabetes, the pancreatic β-cells must secrete additional insulin to compensate for the poor insulin response. Thus, insulin resistance is commonly associated with chronic hyperinsulinemia. The poor response to insulin in insulin resistance is selective for some of insulin's actions. For example, people with insulin resistance have a poor stimulation of glucose transport into muscle and adipose tissue in response to insulin. However, they retain the ability to stimulate lipogenesis in response to insulin. Thus, the hyperinsulinemia associated with insulin resistance can promote excessive lipogenesis. This can lead to increased hepatic triglyceride synthesis and secretion. In Type II diabetes mellitus, the pancreas fails to adequately compensate for insulin resistance. Thus Type II diabetics can still have high insulin levels but not high enough to achieve euglycemia (normal glucose levels) or they can have below-normal insulin levels due to loss of β-cells.

The hypertriglyceridemia of Type II diabetes, unlike that which is found with Type I diabetes, is not due to excessive adipocyte lipolysis. This is because only a small level of insulin action is required to suppress excessive adipose tissue hormone-sensitive lipase activity. In Type II diabetes, there is insufficient adipose tissue lipoprotein lipase and excessive hepatic triglyceride synthesis. Thus, inefficient VLDL triglyceride catabolism and excessive VLDL triglyceride secretion both contribute to the excess VLDL in Type II diabetes.

XVII. TREATMENT OF LIPOPROTEIN DISORDERS

Treatment of lipoprotein disorders is primarily aimed at achieving relatively low VLDL and LDL levels and relatively high HDL levels. Since obesity and insulin resistance are often associated with elevated VLDL, weight loss and exercise are often effective in reducing VLDL. Exercise has an insulin-sensitizing effect on muscle; thus regular exercise can have long-term effects on plasma lipoproteins. Exercise also tends to raise HDL levels. For some people, a reduction in carbohydrate intake (replacing the calories with monounsaturated fat) can lower triglyceride levels, presumably by decreasing the rate of *de novo* lipogenesis in the liver and adipose tissue.

Drugs derived from *fibric acid* (*p*-chlorophenoxyisobutyrate) are widely used to lower triglycerides when diet and exercise fail. These agents increase fatty acid oxidation and decrease VLDL triglyceride secretion. Occasionally, they increase LDL cholesterol and must be used together with an LDL-lowering drug.

Statins are a family of drugs that specifically reduce cholesterol synthesis by inhibiting HMG-CoA reductase.

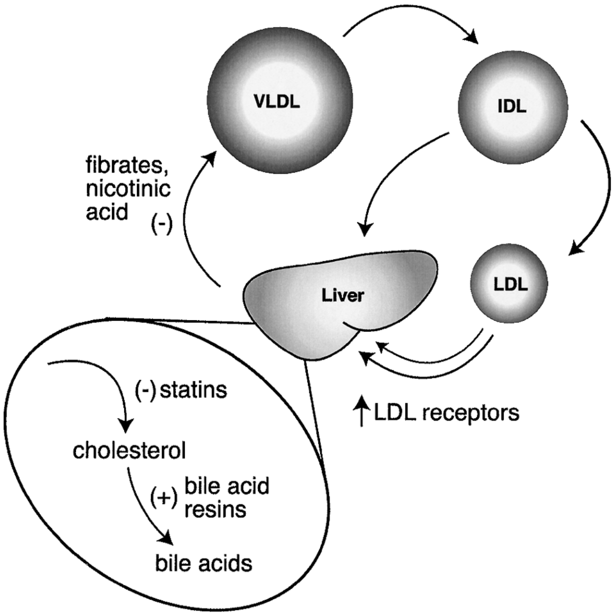


FIGURE 14 Sites of action of three common lipid-lowering drugs. Statins inhibit hepatic cholesterol synthesis. Bile acid sequestrants increase cholesterol catabolism to bile acids. Both agents decrease hepatic cholesterol levels, leading to an upregulation of the LDL receptor. This leads to increased LDL clearance from the circulation. In addition, LDLreceptor upregulation decreases VLDL secretion. Fibrates and nicotinic acid decrease VLDL triglyceride secretion. Agents that lower VLDL tend to raise HDL through the mechanisms described in Fig. 13.

The dearth of cholesterol in the liver leads to upregulation of the LDL receptor (see Fig. 14). In most patients who respond to statins, LDL production is decreased. In addition, there is often an increase in LDL clearance. These drugs are widely used and have made a significant impact on cardiovascular disease in several nations.

Nicotinic acid, an over-the-counter coenzyme, when used at very high doses (1–3 g/day), lowers triglycerides and can achieve a significant increase in HDL. In many individuals, this agent causes an unpleasant skin flushing.

Bile acid sequestrants are charged resins that are ingested in a liquid suspension. They bind to bile acids in the intestine and prevent their reabsorption. Since bile acids normally feed back on their own synthesis from cholesterol, these agents evoke a compensatory increase in bile acid synthesis. The diversion of liver cholesterol for bile acid production leads to an upregulation of the LDL receptor and thus a reduction in LDL levels. Because bile acid sequestrants increase cholesterol catabolism and statins decrease cholesterol synthesis, the two agents together act synergistically.

The majority of people with low HDL have high triglycerides. Their HDL levels usually rise if their triglyceride levels are reduced. However, a significant number of people have low HDL and normal triglycerides; they have

primary hypoalphalipoproteinemia. The treatment options for these individuals are limited. Thus, the next frontier in drug development is to develop treatments for low HDL or ones that enhance the catabolism of cholesterol. The discovery of lipid transport proteins such as ABCA1 provides potential new targets for drug development.

XVIII. FINAL PERSPECTIVE

Cholesterol has played a distinguished role in the history of chemistry, medicine, physiology, and pathology. But, unlike any other biomolecule, cholesterol has also taken center stage as a cultural entity. In most nations, regardless of scientific training, people think about cholesterol when they shop for food, plan their diets, and make lifestyle choices. There is a large food supplement industry that promotes products based on claims to lower cholesterol.

One positive outcome of the cultural presence of cholesterol has been a high level of public education in the area of lipid metabolism—millions of people know the difference between LDL and HDL. This level of public sophistication in details of biochemistry is unprecedented and demonstrates that scientific literacy in other fields is also possible and achievable. During the past 30 years, people in many Western nations have adopted healthier diets and lifestyles, leading to a dramatic drop in the rate of coronary heart disease.

SEE ALSO THE FOLLOWING ARTICLES

Bioenergetics • Enzyme Mechanisms • Gene Expression, Regulation of • Ion Transport Across Biological Membranes • Protein Synthesis

BIBLIOGRAPHY

Ballantyne, C. M., Grundy, S. M., Oberman, A., Kreisberg, R. A., Havel, R. J., Frost, P. H., and Haffner, S. M. (2000). "Hyperlipidemia: Diagnostic and therapeutic perspectives," *J Clin. Endocrinol. Metab.* **85,** 2089–2112.

Bors, P., Zelcer, N., and van Helvoort, A. (2000). "ABC transporters in lipid transport," *Biochim. Biophys. Acta.* **1486,** 128–144.

Brown, M. S., and Goldstein, J. L. (1986). "A receptor-mediated pathway for cholesterol homeostasis," *Science* **232,** 34–47.

Brown, M. S., and Goldstein, J. L. (1999). "A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood," *Proc. Natl. Acad. Sci. USA* **96,** 11041–11048.

Davis, R. A. (1999). "Cell and molecular biology of the assembly and secretion of apolipoprotein B-containing lipoproteins by the liver," *Biochim. Biophys. Acta* **1440,** 1–31.

Ginsberg, H. N. (2000). "Insulin resistance and cardiovascular disease," *J. Clin. Invest.* **106,** 453–458.

Krieger, M. (1999). "Charting the fate of the 'good cholesterol': Identification and characterization of the high-density lipoprotein receptor SR-BI," *Annu. Rev. Biochem.* **68,** 523–558.

Mahley, R. W., and Huang, Y. (1999). "Apolipoprotein E: From atherosclerosis to Alzheimer's disease and beyond," *Curr. Opin. Lipidol.* **10,** 207–217.

Rothblat, G. H., de la Llera-Moya, M., Atger, V., Kellner-Weibel, G., Williams, D. L., and Phillips, M. C. (1999). "Cell cholesterol efflux: Integration of old and new observations provides new insights," *J. Lipid Res.* **40,** 781–796.

Russell, D. W. (1999). "Nuclear orphan receptors control cholesterol catabolism," *Cell* **97,** 539–542.

Tall, A. R., Jiang, X.- C., Luo, Y., and Silver, D. (1999). "George Lyman Duff memorial lecture: Lipid transfer proteins, HDL metabolism, and atherogenesis," *Arterioscler. Thromb. Vasc. Biol.* **20,** 1185–1188.

Terpstra, V., van Amersfoort, E. S., van Velzen, A. G., Kuiper, J., and van Berkel, T. J. (2000). "Hepatic and extrahepatic scavenger receptors: Function in relation to disease," *Arterioscler. Thromb. Vasc. Biol.* **20,** 1860–1872.

Willnow, T. E., Nykjaer, A., and Herz, J. (1999). "Lipoprotein receptors: New roles for ancient proteins," *Nat. Cell Biol.* **1,** E157–E162.

Membrane Structure

Anna Seelig
Joachim Seelig
University of Basel

I. Introduction
II. Membrane Lipids
III. The Membrane–Water Interface
IV. Hydrophobic Core Region
V. Phase Behavior of Lipids and Membrane Domain Formation
VI. Interaction of Membrane Lipids with Amphiphilic Molecules and Transmembrane Proteins
VII. Concluding Remarks

GLOSSARY

Biological membrane A very thin sheath of biological material (thickness ~10 nm to 15 nm) which constitutes the envelope of living cells and also of intracellular organelles, separating them from the environment. Membranes are made up from a lipid bilayer into which proteins are embedded. They are highly organized but are nevertheless fluid enough to allow considerable translational, rotational, and flexing movements of the constituent lipid and protein molecules.

Lipid bilayer A double layer of lipid molecules organized in a tail-to-tail arrangement. It is an anisotropically ordered fluid that has a number of properties in common with smectic liquid crystals. The normal to the surface of the lipid bilayer constitutes an axis of motional averaging. From an optical point of view, bilayer membranes thus behave like uniaxial crystals with the bilayer normal as the optical axis (director axis $\vec{n}$). All fast molecular motions such as rotational and flexing movements are thus characterized, on the average, by a cylindrical symmetry with the bilayer normal as the axis of motional averaging.

Deuterium nuclear magnetic resonance (^{2}H-NMR) By means of chemical synthesis or biochemical incorporation, protons in lipid molecules can be selectively replaced by deuterium atoms. Since the van der Waals radii of the two isotopes are identical, this substitution leaves the membrane virtually unchanged, which is in contrast to other bulkier reporter groups. Deuterium nuclear magnetic resonance provides information on the *order* and *mobility* of the molecule. Structural information comes from the deuterium quadrupole splitting, dynamic information is derived from NMR relaxation times.

Quadrupolar splitting ($\Delta\nu_Q$) The deuterium nucleus has a spin = 1 and, due to its electric quadrupole moment, the anisotropic motion within the membrane will

give rise to a quadrupole splitting, $\Delta\nu_Q$ (kHz). In an unoriented sample, as most membrane preparations are, the deuterium quadrupole interactions give rise to a characteristic powder pattern. The spectrum has two distinct peaks, the separation of which is the so-called deuterium quadrupole splitting, $\Delta\nu_{Q\,powder}$. The deuterium quadrupole splitting may by used to calculate the deuterium order parameter S_{CD} according to

$$\Delta\nu_{Qpowder} = (3/4)(e^2qQ/h)S_{CD}$$

The static deuterium quadrupole coupling constant is $(e^2qQ/h) = 170$ kHz for aliphatic carbon-deuterium (C—D) bonds. A change in the residual quadrupole splitting can be caused by two different mechanisms. First, the angle of the molecular fluctuations may increase or decrease, secondly, the molecule may undergo a conformational change which alters the orientation of the C—D bond vector with respect to the bilayer normal.

Order parameter (S_{CD}) The deuterium order parameter is a measure of the motional anisotropy of the particular C—D bond investigated and yields its time-averaged orientation. If Θ denotes the instantaneous angle between the C—D bond and the direction of the bilayer normal then S_{CD} is defined as

$$S_{CD} = (1/2)(3\,\overline{\cos^2\Theta} - 1)$$

where the bar denotes a time average.

Order parameter (S_{mol}) Assuming an axial symmetry of the segment motion S_{CD} can further be related to the molecular order parameter S_{mol} according to

$$S_{mol} = -2S_{CD}.$$

If the chains are fixed in an all-trans conformation and are just rotating around the long molecular axis, the molecular order parameter would be unity. The other extreme is that of a completely statistical movement through all angles of space, leading to $S_{mol} = 0$. This simple statistical interpretation of S_{CD} is not possible if specific geometric effects come into play as, for example, in the case of the *cis*-double bond.

Order profile of the lipid bilayer It shows the variation of the order parameter, S_{mol} or S_{CD}, with the position of the segment in the chain and is an expression of the average angular fluctuations around the bilayer normal.

Spin-lattice relaxation time (T_1) The spin lattice relaxation time depends on both the *ordering* (S_{CD}) and the *rate of motion* (correlation time, τ_C). Assuming a motion sufficiently characterized by a single correlation time, τ_C, the following expression holds for the short correlation time limit:

$$1/T_1 = \frac{3}{8}\left(\frac{e^2qQ}{\hbar}\right)^2\left(1 - S_{CD}^2\right)\tau_C$$

Phosphorous nuclear magnetic resonance (^{31}P-NMR) No isotope labeling is required for ^{31}P-NMR spectroscopy. The chemical shielding anisotropy, $\Delta\sigma$, in ^{31}P-NMR is comparable to the deuterium quadrupole splitting in ^{2}H-NMR and can be determined from the edges of the spectrum.

THE MAIN STRUCTURAL element of the biological membrane is the lipid bilayer. Lipid molecules, when brought into contact with water, spontaneously organize themselves into a bilayer leaflet: The polar lipid headgroups remain in the aqueous environment while the fatty acid tails form the inner hydrophobic core. The lipid bilayer is thus a "sandwich"-like structure with the polar group as the "bread" and the fatty acyl chain as the "butter." The structure of the lipid bilayer and the interaction of the lipid molecules with their environment, such as metal ions, peptides, and proteins, are the themes presented here. Using solid-state nuclear magnetic resonance (NMR) techniques, a quantitative analysis of the molecular ordering and dynamics of a lipid bilayer has become possible with a segment-to-segment resolution. Lipid bilayers—and also intact biological membranes—are not rigid but can be classified in physical terms as smectic liquid crystals. The lipids within each bilayer undergo rapid translational and rotational motion. The packing of the hydrocarbon chains is best described in terms of statistical order profiles. In contrast, well-defined conformations are observed for the glycerol backbone and to some extent also for the polar head groups. Both the order profile and the orientation of the polar groups can vary considerably depending on the external conditions and constitute regulatory elements for the function of the biological membrane.

I. INTRODUCTION

Biological membranes segregate cells and organelles, act as barriers for the passive transport of matter, and support a wide range of important metabolic processes, including active transport, energy flow, signal transduction, and motility. The two main components of membranes are lipids and proteins. Depending on the type of membrane, lipids contribute between 20 and 80% by weight to the total membrane mass, the rest being protein. The lipid molecules are predominantly arranged in a bilayer structure with the hydrophilic head groups facing the aqueous environments and the fatty acyl chains forming the inner hydrophobic core. Minor but functionally important components of membranes are carbohydrates. They are covalently attached to either lipids (glycolipids) or proteins (glycoproteins) and are restricted to the outer leaflet of the bilayer membrane.

The distribution of the lipids between the inner and outer leaflet of a biological membrane is asymmetric, with the outer surface being enriched in phosphatidylcholine (PC) and the inner, cytosolic surface in phosphatidylethanolamine (PE) and phosphatidylserine (PS). As a result, the outer lipid membrane surface is electrically neutral and the inner negatively charged. Spontaneous randomization (flip-flop) of zwitterionic lipids between the two leaflets is extremely slow. Specific transport proteins, belonging to the adenosine triphosphate (ATP) binding cassette, further maintain lipid asymmetry. Lipid asymmetry may play a role for the proper orientation of membrane proteins.

The lipid composition of natural membranes is quite heterogeneous. The large variability with respect to head groups, chain length, and extent of *cis*-unsaturation results in thousands of chemically different lipids. For example, the membrane of the red blood cell contains about 400 chemically different lipids. The lipid composition and the lipid-to-protein ratio of a given membrane are relatively well defined, suggesting a correlation between lipid composition and membrane function. At present, little is known of how the lipid composition is controlled and why biological membranes contain so many different lipids. Many biological membranes can adapt to changing external conditions such as temperature or long-term exposure to drugs or alcohol by modifying their lipid composition in order to maintain the optimal conditions for cell growth.

Hydrophobic membrane proteins and lipids are difficult to crystallize compared to water-soluble biological molecules. Consequently, structural information on membrane components has become available at a much slower pace than on water-soluble proteins or DNA.

The situation is even worse for membrane lipids. Not a single, naturally occurring phospholipid with unsaturated hydrocarbon chains has yet been crystallized. However, nearly 40 crystal structures of closely related synthetic glycerolipids with saturated hydrocarbon chains have been solved by X-ray. On the structural level, little is known about the interactions of proteins with lipid bilayer environments. Detergent molecules have been detected in some of the X-ray structures, and a small number of studies discuss lipids bound to proteins. An example is cytochrome C oxidase crystals, where the lipids were found to be arranged in a bilayer structure.

Magnetic resonance techniques, in particular phosphorus (^{31}P) and deuterium (^{2}H) magnetic resonance, in combination with selectively deuterated lipids, have yielded quantitative information on the ordering, motional anisotropy, and dynamics of membrane components. This information is essential for understanding the function of biological membranes.

The different structural elements of the lipid membrane include the polar part, constituting the interface to the aqueous compartment and consisting of the head group proper, the glycerol backbone, and the ester linkages. The molecular details of the membrane surface, including the electric surface charges, are relevant for membrane recognition by molecules such as enzymes dissolved in the extra- and intracellular space. The hydrophobic core region of membranes is formed by the fatty acyl chains of lipids. The order and dynamics of the hydrophobic core determine the permeability of the membrane to molecules such as drugs and may modulate the function of transmembrane proteins. In addition to these elelments, we will also discuss the interaction of the lipid membrane with amphiphilic molecules, which penetrate into the hydrophobic core region, and with intrinsic membrane proteins. The NMR results obtained by solid-state NMR will be compared with those obtained with neutron and X-ray diffraction and with recent molecular dynamics simulations of membranes.

II. MEMBRANE LIPIDS

Naturally occurring lipids can be divided essentially in two groups: (1) phospholipids containing glycerophosphate as the anchor group for fatty acids, and (2) lipids containing backbones other than glycerol. Phospholipids may be further subdivided into nitrogen-containing lipids, such as phosphatidylcholine (PC), -ethanolamine (PE), -serine (PS), and plasmalogens, and nitrogen-lacking lipids, such as phosphatidic acid (PA), phosphatidylglycerols (PG), cardiolipin (CL), and phosphoinositols (PI). Examples for lipids not containing the glycerol backbone are the sphingolipids and glycosphingolipids, derived from sphingosine or dihydrosphingosine and are mainly found in nerve cells and in brain. Sphingolipids comprise ceramides and sphingomyelins, whereas glycosphingolipids can be subdivided into cerebrosides and gangliosides, both bearing carbohydrate head groups as characteristic structural elements. An important lipid component in eukaryotic (but not in prokaryotic) cells are the sterols with cholesterol as the principal representative.

The lipids in eukaryotic cell membranes are mainly nitrogen-containing phospholipids. They are involved in the maintenance of the barrier properties of membranes and provide the optimal conditions for transmembrane protein functioning. Some phospholipids also play a decisive role in cell signaling processes. Phosphatidic acid was shown to enhance the membrane binding of phospholipase C-β_1 and to stimulate hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2), resulting in the formation of diacylglycerol and inositol 1,4,5-triphosphate. The released diacylglycerols may then further activate kinases.

Sphingolipids and glycosphingolipids are believed to be structural as well as signaling constituents of membranes.

Sphingomyelin hydrolysis yields ceramide, a lipid mediator involved in regulating cell growth, cell differentiation, and cell death. Glycosphingolipids act as specific recognition sites in eukaryotic cells, and they determine blood-group, organ, and tissue specificity and are further involved in tissue immunity and cell–cell recognition.

The phospholipids found in prokaryotic (bacterial) and eukaryotic (mammalian) cell membranes usually contain saturated as well as *cis*-unsaturated fatty acyl chains, the most abundant being the saturated palmitoyl (C 16:0) and the *cis*-unsaturated oleoyl chains (C 18:1, *cis*). There is a strong positional preference for the two types of fatty acids, with the saturated and the unsaturated chain being localized at positions 1 and 2, respectively, of the glycerol backbone of the lipid molecule. Phospholipids with a single *cis*-double bond are predominant, but lipids containing more than one double bond also occur quite commonly. In membranes of the nervous system, polyunsaturated fatty acids appear to be critical for proper membrane functioning. The length of the fatty acyl chains and the degree of chain unsaturation as well as the size, the charge, and the hydrogen-bonding capacity of headgroups determine the intermolecular lipid–lipid interactions reflected in the lipid packing density and the gel-to-liquid crystal phase transition temperatures of lipid membranes. The effect of headgroups on the gel-to-liquid crystal phase transition temperature, T_c, is illustrated by the following series of lipids (C 16:0) mixed with water: PC ($T_c = 41°C$) ~ PG ($T_c = 41°C$) ~ SM ($T_c = 41°C$) < PS deprotonated ($T_c = 54°C$) < PS protonated ($T_c = 62°C$) < PE ($T_c = 64°C$) < PA ($T_c = 71°C$) < GalCer ($T_c = 82°C$). Knowledge of the gel-to-liquid crystal phase transitions is of relevance for the question of lipid domain formation to be discussed below.

III. THE MEMBRANE–WATER INTERFACE

The membrane–water interface comprises the lipid headgroup proper, with tightly bound hydration water and a more loosely packed extended hydration layer; the glycerol backbone; and the ester linkages of the fatty acyl chains. Due to conformational constraints, the carbonyl group of the *sn*-2 chain is particularly close to the lipid–water interface, while that of the *sn*-1 chain is inserted deeper into the membrane interior.

A. Headgroup and Glycerol Backbone Conformation of Phospholipids

The crystal structures of three synthetic lipids are shown in Fig. 1. Essential elements of these crystal structures are carried over into the liquid crystalline state as revealed by solid-state NMR. The main features are as follows: For PC, PE, and PG, the glycerol backbone is oriented perpendicular to the bilayer surface, while the polar headgroups are almost parallel to the membrane surface. Neutron scattering experiments of selectively deuterated lipid headgroups in liquid crystalline and gel state membranes determine the mean label position with an accuracy of up to ±1 Å and provide independent support for the almost parallel headgroup orientation of PC, PE, and PG.

The headgroup orientations of PC, PE, and PG bilayers in the liquid crystalline phase, in the gel phase, and in single crystals are thus very similar and independent of the dynamic state of the membrane. The correlation times of the segmental and collective motions of the head groups decrease abruptly by more than two orders of magnitude at the gel-to-liquid phase transition; nevertheless, the average conformation remains unaltered.

Phosphatidylserine measured at neutral pH and in the absence of ions is similar to the other phospholipids with respect to the glycerol backbone, but differs distinctly in its headgroup orientation and motion. The PS headgroup is rigid and exhibits little internal flexibility. A crystal structure is not available so far.

For the comparison of NMR and X-ray diffraction measurements, the effect of membrane hydration can be relevant. A minimum of 11 to 16 water molecules per lipid molecule is needed to form a primary hydration shell for PC, PE, and PG. Additional water is in exchange with the primary hydration shell. With increasing hydration (10–70 wt% H_2O) the $-P^-$-N^+ dipole of the phosphocholine headgroup was shown to move with its cationic end away from the hydrocarbon layer. This explains why the $-P^-$-N^+ dipoles in liquid crystalline membranes are generally slightly tilted away from the membrane surface up to an angle of about 30°, while they are oriented parallel to the surface in the crystal structure.

B. Ester Linkage of the sn-2 Fatty Acyl Chain Is Part of the Lipid–Water Interface

If the two fatty acyl chains of lipids are deuterated at methylene groups immediately next to the ester linkage they give rise to quite different quadrupole splittings, indicating that the beginnings of the two chains have different conformations. The inequivalence of the two chains was first observed for 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) in its liquid crystalline phase. It is also preserved in the presence of transmembrane proteins in reconstituted and in natural membranes.

The conformational difference between the two fatty acyl chains is particularly pronounced for the C-2 segments (i.e., the segment next to the ester linkage) which give rise to three separate resonance splittings (cf.

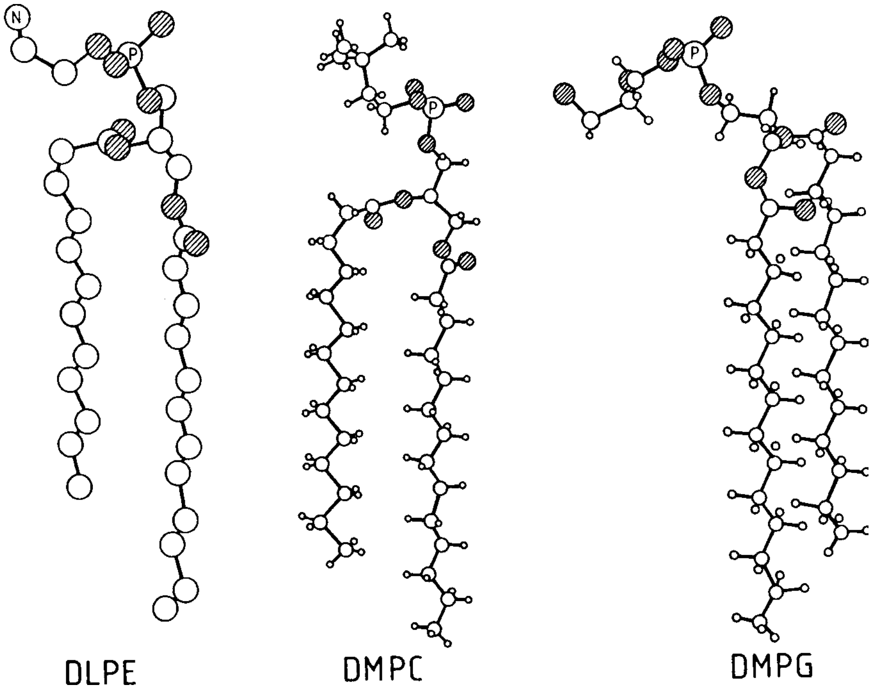


FIGURE 1 Single-crystal structures of three phospholipids. The lipids are 1,2-dilauroyl-*sn*-glycero-3-phosphoethanolamine (DLPE) [Hitchcock et al. (1974). *Proc. Natl. Acad. Sci. USA* **71,** 3036], 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) [Pearson and Pascher (1979). *Nature* **281,** 49], and 1,2-dimyristoyl-*sn*-glycero-3-phosphogycerol (DMPG) [Pascher et al. (1987). *Biochim. Biophys. Acta* **896,** 77]. Structural features which are carried over into liquid-crystalline membranes: (1) the polar groups are oriented at approximately a right angle to the hydrocarbon chains, and (2) in DLPE and DMPC the *sn*-2 fatty acid chain is bent at the C-2 segment while the *sn*-1 chain is straight. A bent *sn*-2 chain is a common property of phospholipids in biomembranes. Only one of two possible conformations is shown for each lipid [Seelig et al. (1987). *Biochemistry* **26,** 7535].

Fig. 2A). It is still detectable at position C-3 (Fig. 2C) but averages out at label positions deeper in the hydrocarbon layer (Fig. 2D, the C-10 segment).

At a molecular level, the different splittings indicate that the *sn*-1 chain extends perpendicularly to the bilayer surface with all segments, while the *sn*-2 chain starts out parallel to the bilayer surface and makes a 90° bend after the C-2 segment in order to keep the *sn*-1 and *sn*-2 chains parallel to each other in agreement with the crystal structure shown for PC and PE in Fig. 1. Furthermore, labeling of the glycerol backbone suggests the possibility of two long-lived conformations of the glycerol constituent.

The conformation of the two fatty acyl chains near the glycerol moiety was further investigated by synthesizing specifically deuterated 1,3-dipalmitoyl-*sn*-glycero-2-phosphocholine. As seen in Fig. 2B the spectrum of the liquid crystalline phase of *sn*-glycero-2-phosphatidylcholine deuterated at position C-2 shows only a single quadrupolar splitting, indicating that the C-2 segments of *both* fatty acyl chains are now aligned parallel to the bilayer surface and the chains are bent perpendicularly to the surface only after the C-2 segment.

One of the most extensively investigated proteins active at the membrane surface is phospholipase A_2. This enzyme is water soluble, attacks the membrane from the aqueous phase, and acts specifically on the *sn*-2 chain. In light of the conformational results presented here, the attack of phospholipase A_2 is facilitated by the orientation of the *sn*-2 ester bond parallel to the membrane surface.

C. Phospholipid Headgroup Response to Ions

The quadrupolar splitting, $\Delta\nu_Q$, of headgroup-deuterated PC or PE varies linearly as a function of the total amount of electric surface charge, and changes in opposite directions are induced by positive and negative surface charges. This indicates that the phosphocholine and the phosphoethanolamine dipoles are sensitive to electric charges at the membrane surface and function as an electrometer. As an example, the interaction of α-CD_2-POPC with an anionic and a cationic amphiphile is shown in Fig. 3.

The chemical nature of the ion imparting the membrane surface charge appears to be of secondary importance. A variety of chemically different, charged compounds including metal ions, local anesthetics, peptides,

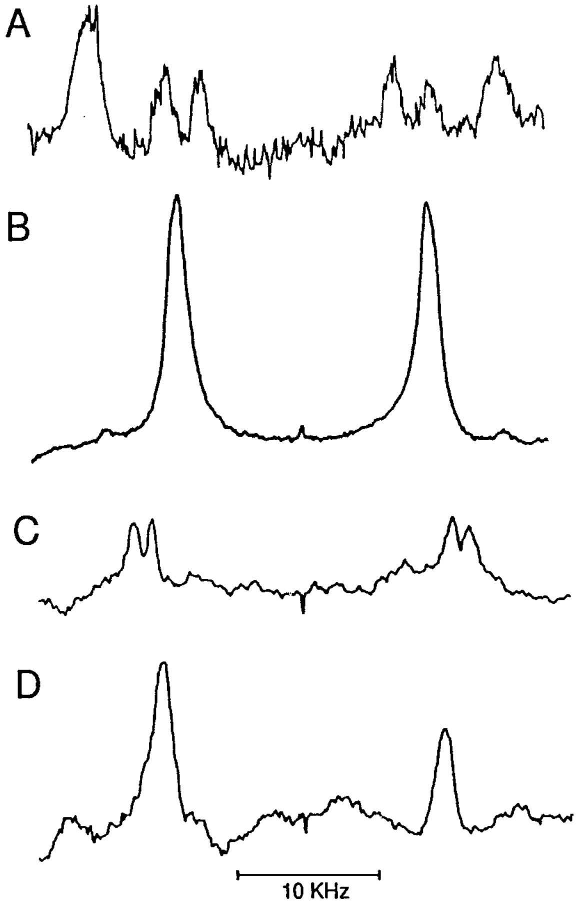


FIGURE 2 Deuterium magnetic resonance spectra of *sn*-2 and *sn*-3 phosphatidylcholine bilayers deuterated at different positions (50 wt% lipid, 50 wt% H_2O). (A) 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine deuterated in both chains at the C-2′ segment [Seelig and Seelig (1975). *Biochim. Biophys. Acta* **406,** 1]; (B) 1,3-*bis*-([2′,2′-$^2H_2$]palmitoyl)-*sn*-glycero-2-phosphocholine [Seelig et al. (1980). *Biochemistry* **19,** 2215); (C) 1,2-dipalmitoyl-*sn*-glycero 3-phosphocholine deuterated in both chains at the C-3′ segment; (D) 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine deuterated in both chains at the C-10′ segment [Seelig and Seelig (1974). *Biochemistry* **13,** 4839].

hydrophobic cations and anions, amphiphiles, and lipids have been shown to yield similar results when incorporated as guest molecules into PC or PE membranes. This is demonstrated in Fig. 4, which summarizes the experimental results in a rather condensed representation. If a guest molecule is added to a headgroup deuterated phospholipid, the α- and β-quadrupole splittings ($\Delta\nu_\alpha$, $\Delta\nu_\beta$) change linearly with increasing concentration. In Fig. 4 the slopes, m_α and m_β, of such $\Delta\nu_\alpha$ and $\Delta\nu_\beta$ versus concentration plots are shown. A linear correlation exists between m_α and m_β, with a slope of -0.5 for cations and -1 for anions. The molecular interpretation is as follows: A positive electric charge on the membrane surface moves the N^+ end of the headgroup $-P\text{-}N^+$ dipole away from the membrane surface, and a negative charge moves the N^+ end towards the hydrocarbon phase. The out-of-plane movement of the phospholipid headgroup dipole creates a local electric field across the membrane, which can easily reach a field strength of 10^5 V/cm. Such high electric fields can, in principle, entail conformational changes of membrane-bound proteins, and the lipid dipole field could thus play a regulatory role in membrane function.

If the membrane contains negatively charged lipids to begin with, the concentration of cationic compounds at the membrane surface is drastically enhanced, facilitating the binding and also providing an additional mechanism of electric modulation.

D. Headgroup Orientation in Glycolipids and Glycosphingolipids and Their Influence on Phospholipid Headgroups

The deuterium order parameter of headgroup-labeled glycolipids and glycosphingolipids generally show a headgroup orientation in which the sugar residues project essentially straight up from the bilayer surface into the aqueous region, permitting maximum hydration of the glucose hydroxyl groups by water. The glucosyl headgroup appears to be rather rigid, but rotates with a rotational diffusion constant of $\sim 10^8$ s^{-1}.

The headgroup conformational changes of deuterium-labeled PC observed in the presence of glycolipids and glycosphingolipids were shown to be qualitatively similar to those of negatively charged ions (cf. Fig. 4). However, in comparison to the effects induced by charged substances, these effects were modest.

IV. HYDROPHOBIC CORE REGION

A. Fatty Acyl Chain Order in Saturated Lipid Membranes

The hydrocarbon chains of the lipid bilayer are in a liquid-like state as evidenced by X-ray diffraction, electron spin resonance spectroscopy, and differential scanning calorimetry studies. A quantitative characterization of the hydrocarbon chain order in lipid bilayers by means of ^{2}H-NMR became possible by selectively deuterating both fatty acyl chains in a lipid molecule. Measurement of the deuterium quadrupole splittings, $\Delta\nu_Q$, allowed calculation of the order parameter of the C—D bond vector at each labeled carbon atom. The variation of the order parameter $|S_{CD}|$ with the position of the labeled carbon atom in the membrane is the so-called "order profile." An

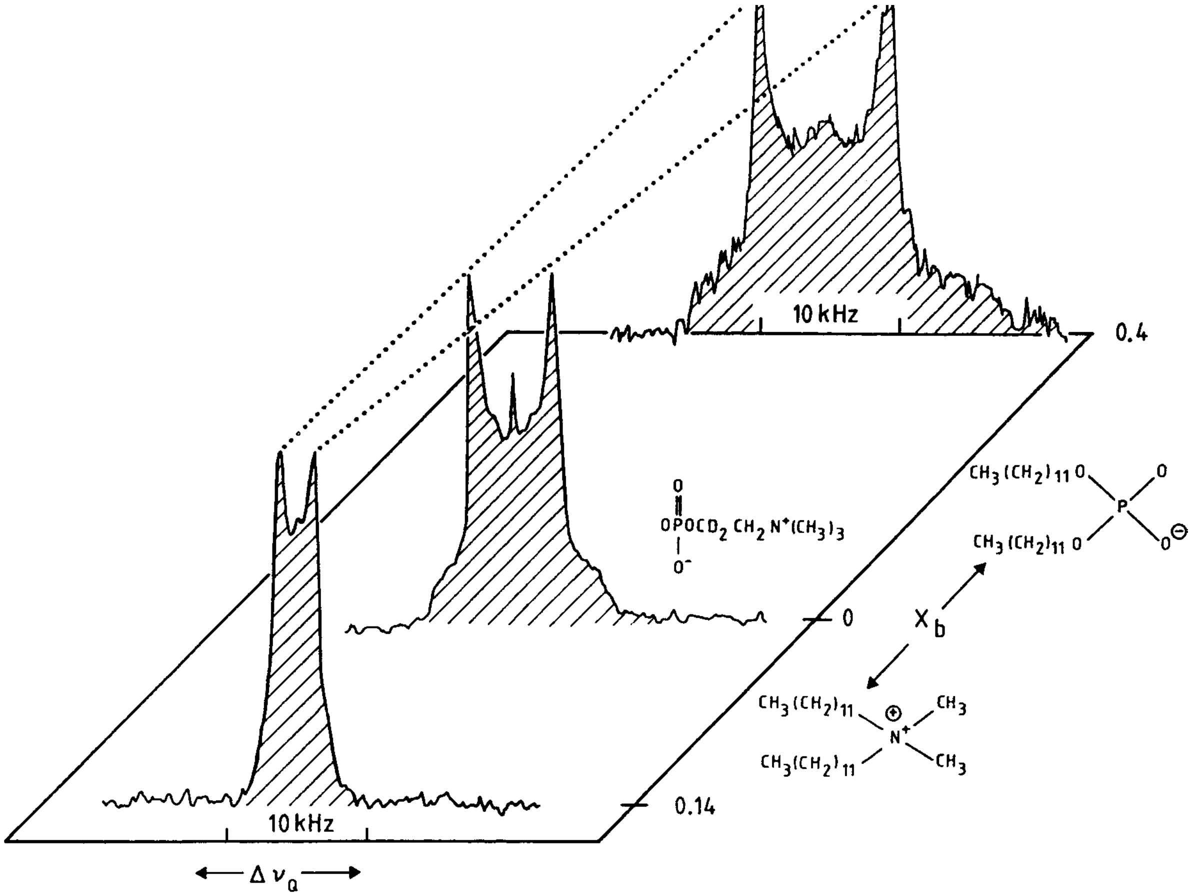


FIGURE 3 Charged amphiphiles in lecithin membranes induce a reorientation of the −P-N$^+$ dipole. The figure shows deuterium NMR spectra of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine deuterated at the choline headgroup (α-CD_2-POPC) without amphiphile ($X_b = 0$), with cationic amphiphile ($X_b = 0.14$), and with anionic amphiphile ($X_b = 0.4$). Positive charges decrease the quadrupole splitting of the α-segment; negative charges increase it [Seelig et al. (1987). *Biochemistry* **26,** 7535].

example is shown in Fig. 5 for a membrane composed of DPPC at temperatures $T > 41°C$ (liquid crystalline phase). The segmental order parameters are approximately constant for the first nine chain segments, but decrease towards the central part of the bilayer. The chain ordering can be explained on the basis of the rotational isomeric model for hydrocarbon chains. In the region of constant order parameters, *trans*-gauche isomerizations occur only in complementary pairs, leaving the hydrocarbon chains essentially parallel to each other. This leads to well-ordered bilayers with disordered hydrocarbon chains. The decrease of the order parameter in the central region is due to increasing contributions of gauche states. The total number of gauche isomeric states was estimated to be between 3 and 6 per chain. The quantitative evaluation of the deuterium data further yields the thickness of the hydrocarbon region of the bilayer. For DPPC bilayers in the liquid crystalline phase, an average thickness of about 30 Å and a thermal expansion coefficient of $-2.5\ 10^{-3}$/K was derived, in good agreement with X-ray diffraction experiments. It should be noted that this approach is also valid for determining the bilayer thickness of highly unsaturated membranes. The DPPC order parameter profile (Fig. 5) serves as a "gold standard" for molecular dynamics simulations of bilayers. Similar order profiles have now been established for a large variety of pure lipid membranes as well as intact biological membranes.

B. Incorporation of *cis*-Double Bonds

As mentioned before, almost all biological lipids contain unsaturated fatty acyl chains at the *sn*-2 position. The incorporation of the *cis*-double bond introduces a kink in the otherwise straight chain and reduces the gel-to-liquid crystal phase transition temperature by some

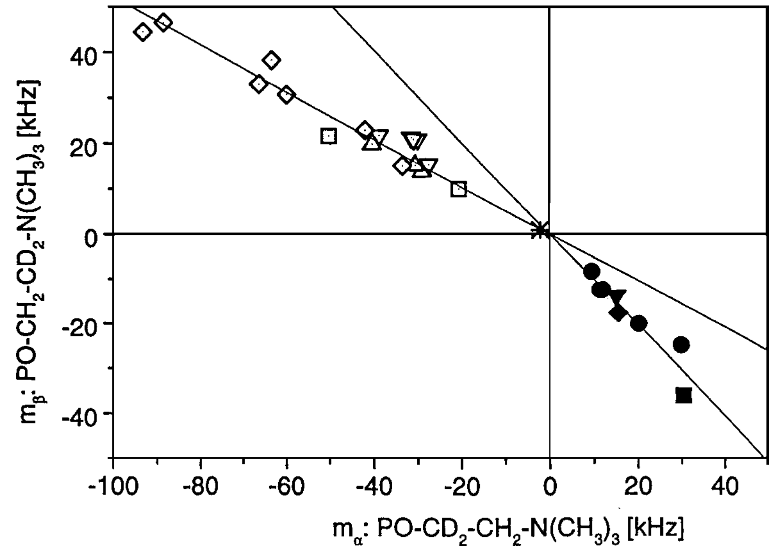


FIGURE 4 Variation of the quadrupole splittings of the α- and the β-segment in the headgroup of phosphatidylcholine bilayers upon binding of cationic (open symbols) and anionic (solid symbols) substances. The difference between the quadrupole splittings of the α-segment with, $(\Delta\nu(X_b))$, and without $(\Delta\nu_i^\circ)$ guest molecule is plotted versus that of the β-segment under identical conditions, $(\Delta\nu(X_b) - \Delta\nu_i^\circ)/X_b = m_i$, where X_b, is the mole fraction of bound guest molecule and *i* stands for α or β. Metal ions (□) [Altenbach and Seelig (1984). *Biochemistry* **23,** 3913; Macdonald and Seelig (1987). *Biochemistry* **26,** 1231, 6292]; drugs (△) [Boulanger *et al.* (1981). *Biochemistry* **20,** 6824; Seelig *et al.* (1988). *Biochim. Biophys. Acta* **939,** 267; Bauerle and Seelig (1991). *Biochemistry* **30,** 7203]; amphiphilies (▽) [Scherer and Seelig (1989). *Biochemistry* **28,** 7720]; peptides (◇) [Beschiaschvili and Seelig (1991). *Biochim. Biophys. Acta* **1061,** 78; Kuchinka and Seelig (1989). *Biochemistry* **28,** 4216; Roux *et al.* (1989). *Biochemistry* **28,** 2313; Spuhler *et al.* (1994). *J. Biol. Chem.* **269,** 23904; Wieprecht *et al.* (2000). *Biochemistry* **39,** 442; Schote *et al.* (2000). *Pharm. Res.*]; electrically neutral detergent (∗) [Wenk and Seelig (1997). *Biophys. J.* **73,** 2565]; inorganic anion (■) [Macdonald and Seelig (1988). *Biochemistry* **27,** 6769]; peptides (◆) [Schote *et al.* (2000). *Pharm. Res.*]; amphiphiles (▼) [Scherer and Seelig (1989). *Biochemistry* **28,** 7720]; phospholipids (●) [Marassi and Macdonald (1992). *Biochemistry* **31,** 10031; Scherer and Seelig (1989). *Biochemistry* **28,** 7720; Pinheiro *et al.* (1994). *Biochemistry* **33,** 4896]. The slope is characteristic for the sign of the electric charge and is $m = -0.52 \pm 0.01$ for cations and $m = -1.01 \pm 0.05$ for anions [cf. Scherer and Seelig (1989). *Biochemistry* **28,** 7720].

40°C compared to the saturated lipid. The influence of the *cis*-double bond was investigated in bilayers composed of 1-palmitoyl-2-oleoyl-3-*sn*-phosphocholine (POPC) by either labeling the saturated palmitoyl chain or synthesizing deuterated *cis*-unsaturated oleic acid. As seen in Fig. 5 and Fig. 6, the shape of the order profile of the palmitic acyl chain is similar to that observed for the fully saturated DPPC (Fig. 5), but the magnitude of the order

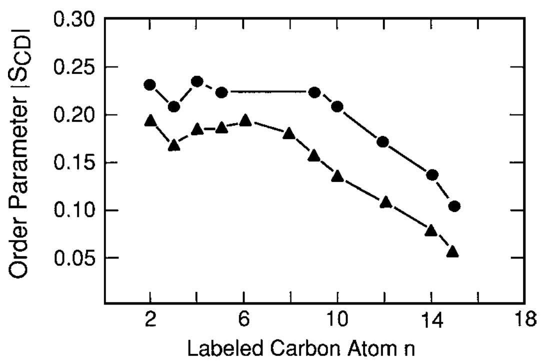


FIGURE 5 Order parameters $|S_{CD}|$ as a function of the labeled carbon atom for 1,2-dipalmitoyl-3-*sn*-phosphocholine (●) and for 1-palmitoyl-2-oleoyl-3-*sn*-phosphocholine (▲) at 42°C. The *sn*-1 chains are labeled. [From Seelig and Seelig (1977). *Biochemistry* **16,** 45.]

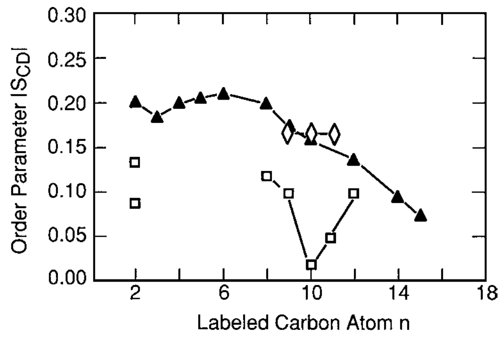


FIGURE 6 The effect of a *cis*- and a *trans*-double bond on the order parameter profile. Bilayers of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine labeled at different positions in the *sn*-1 (▲) and *sn*-2 chain (❑) measured at 27°C; 1-palmitoyl-2-elaidoyl-*sn*-glycero-3- phosphocholine labeled in the *sn*-2 chain (◇) measured at 40°C. [From Seelig and Waespe-Sarcevic (1978). *Biochemisty* **17,** 3310.]

parameters is distinctly smaller in the unsaturated system. This demonstrates that the presence of a *cis*-double bond causes a more disordered conformation of the hydrocarbon chains. Considering the relative flexibility within the palmitic acyl chain, the deuterium resonance data indicate a local stiffening of those segments which are located in the vicinity of the rigid *cis*-double bond. An increase in temperature leads to a further decrease of the order parameters.

The ^{2}H-NMR spectrum of POPC membranes deuterated at the C-9, and C-10 positions of the oleic acyl chain shows two quadrupolar splittings, the larger corresponding to the C-9 deuteron and the smaller to the C-10 deuteron (Fig. 6). The presence of two quadroupole splittings at the same rigid segments is caused by a tilting of the *cis*-double bond with respect to the bilayer normal which produces different orientations for the C–^{2}H vectors of the 9- and 10-carbon atoms. The angle between the bilayer normal and the C=C bond vector was found to be 7 to 8°. The order parameter of 1-palmitoyl-2-elaidoyl-*sn*-glycero-3-choline deuterated at the C-9 and C-10 *trans*-double bond of the elaidic acid chain is also included in Fig. 6. Due to the symmetry of the *trans*-double bond the two C—D vectors at the C-9 and C-10 position make the same angle with the C=C vector axis. They give rise to the same quadrupole splitting, and the evaluation of the order parameter of the C=C axis is straightforward. Taking into account the different geometries, the molecular ordering and the angular fluctuations of the *cis*- and *trans*-double bonds are identical. In addition, there are no quantitative differences between *sn*-1 and *sn*-2 chain segments at this position in the bilayer. The segmental fluctuations around the bilayer normal thus only depend on the distance from the lipid–water interface but not on the specific segment geometry.

The effect of *cis*-unsaturation was also investigated for the glycosphingolipid *N*-(oleoyl-d33)galactosylceramide incorporated at low concentration into liquid crystalline liposomes composed of 1,2-dimyristocyl-3-*sn*-phosphatidylcholine (DMPC) and POPC using the perdeuterated oleoyl chain as the reporter element. The primary effect of *cis*-9,-10 unsaturation in glycosphingolipids proved to be similar to that of *cis*-unsaturation in glycerolipids. It was further shown that the overall dynamics of *N*-(oleoyl)galactosylceramide in fluid phospholipid membranes was very similar to that of glycerolipids with comparable acyl chains.

Increasing *sn*-2 unsaturation from one to six double bonds in PC leads to an inhomogeneous disordering along the neighboring perdeuterated *sn*-1 chain. As a consequence, the effect of a temperature increase leading to a decrease in the average chain length is somewhat less pronounced in lipids with three or more double bonds in the *sn*-2 chain than in lipids with only one double bond.

C. Effect of Cholesterol on the Order and Motion of the Lipid Hydrocarbon Chains

The influence of cholesterol on the order and mobility of lipid bilayers was investigated with both selectively deuterated lipids and deuterated cholesterol. Addition of 50% cholesterol to DPPC and DMPC bilayers was shown to lead to an almost twofold increase of the quadrupole splitting of the labeled fatty acyl chain segment compared to that of a cholesterol-free bilayer. When [3-^{2}H] cholesterol was added to a nondeuterated DPPC bilayer, again a large quadrupole splitting of the cholesterol probe was observed. Both probes lead to the conclusion that a high concentration of cholesterol induces an essentially all *trans*-conformation in those hydrocarbon chain segments which are in contact with the rigid steroid frame. This condensing effect of cholesterol was also observed in monolayer- and neutron-diffraction experiments.

The effect of cholesterol on the order parameter profile of individual fatty acyl chains in DPPC bilayers was simulated by means of molecular dynamics calculations as displayed in Fig. 7. A distinct plateau with an order parameter of the C—D bond vector of $S_{CD} = -0.4$ was detected. This means that the hydrocarbon chains are almost fully extended and that the order parameter of the long molecular axis, $S_{mol} = -2\ S_{CD}$, approaches its maximum value of $S_{mol} = 1$.

In highly unsaturated lipid mixtures, typical for nerve and retinal membranes, cholesterol induces an increase in the order of both saturated and polyunsaturated hydrocarbon chains. However, the increase in order is about a factor of 2 smaller for polyunsaturated than for monounsaturated lipids.

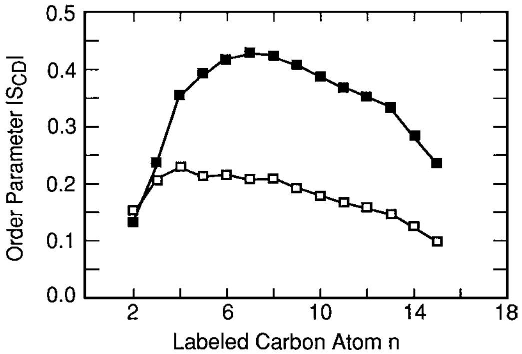


FIGURE 7 Effect of cholesterol. Order parameter of the *sn*-2 chain in DPPC bilayers without (□) and with (■) 50 mol% cholesterol as function of carbon atom. [From Smondyrev and Berkowitz (1999). *Biophys. J.* **77,** 2075.]

As far as lipid headgroups are compared, addition of cholesterol increases the chain order in the sequence 18:0–18:1 PS < 18:0–18:1 PC < 18:0–18:1 PE for the monounsaturated lipid mixture and in the sequence 18:0–22:6 PS < 18:0–22:6 PE ≪ 18:0–22:6 PC for polyunsaturated mixtures. The cholesterol-induced variation of order parameters as a function of the chemical nature of the lipid species suggests a cholesterol-induced formation of lipid microdomains with a headgroup- and fatty-acyl-chain-dependent lipid composition. In particular, under physiological conditions, the formation of PC-enriched microdomains has been proposed in which the saturated *sn*-1 chain is preferentially oriented toward the cholesterol molecule. The lifetime of a lipid molecule in a given cluster, however, is less than 10^{-4} s, and the cluster radius is probably smaller than 25 nm.

In a natural membrane the effect of cholesterol is very similar as in model membranes. This was shown, for example, for *Acholeplasma laidlawii* membranes by the incorporation of perdeuterated and selectively deuterated fatty acids.

V. PHASE BEHAVIOR OF LIPIDS AND MEMBRANE DOMAIN FORMATION

A large number of phase diagrams for binary mixtures combining cholesterol with different saturated and unsaturated phosphatidylcholines have been established. Cholesterol at different bilayer concentrations can promote or suppress lateral segregation of phospholipids of differing acyl chain length.

Addition of 50 mol% cholesterol to selectively deuterated DPPC bilayers leads to an elimination of the gel-to-liquid crystal phase transition at 41°C. In contrast, cholesterol is also found to enhance the tendency of the PC components to exhibit lateral segregation. These seemingly contradictory effects of cholesterol can be readily explained in light of the cholesterol–phospholipid phase diagrams.

The effect of cholesterol on the thermotropic phase behavior of aqueous dispersions of different lipids has been extensively investigated by means of differential scanning calorimetry. The results show an inverse correlation between the strength of intermolecular phospholipid–phospholipid interactions, as manifested by the gel-to-liquid crystalline phase transition temperatures of the pure phospholipids, and the miscibility of cholesterol with the respective bilayer (particularly gel-state bilayers). The miscibility of cholesterol with lipids carrying identical fatty acyl chains decreases in the order: PC ∼ PG ∼ SM > PS > PE > diglucosyl- and monoglucosyl-diacylglycerol > GalCer. However, if the higher melting components are dispersed as *minor* components of total lipid in a host matrix consisting of, for example, 1-stearoyl-2-oleoyl-phosphatidylcholine and cholesterol, neither short-chain nor long-chain cerebrosides or sphingomyelins show phase separation in the physiological temperature range despite their high phase transition temperatures.

Mixtures of cholesterol and sphingolipids have recently attracted attention since spingolipid–cholesterol domain formation has been observed in mammalian cell membranes upon cooling to 4°C and extraction with Triton X-100. This phenomenon has also been termed "lipid raft" formation. Lipid rafts exhibit a high lateral packing density and are suggested to entail a sorting of GPI-anchored proteins. The bulky intrinsic proteins remain in the fluid phase. At room temperature, lipid rafts are no longer detectable. Nevertheless, they are assumed to prevail as microdomains at growth temperature and to be relevant for membrane trafficking and protein sorting in mammalian cells.

Although domain formation is now a common theme among biologists, an unambiguous physical–chemical characterization of domains under physiological conditions is still missing. On physical grounds, domain formation is most likely to occur in mixtures of lipids with widely different gel-to-liquid crystal phase transition temperatures. Phase separation will occur if the measuring temperature is below the phase transition temperature of one of the components of the mixture and if this component constitutes a major lipid fraction. Lipids exhibiting high phase transition temperatures generally have long saturated acyl chains and small headgroups, or headgroups that may interact via hydrogen bonding. Typical examples are sphingolipids, glycosphingolipids, or long-chain phosphatidylethanolamines.

As a further mechanism, electrostatic interactions of anionic lipids with cationic compounds may also induce domain formation. Due to the biochemical complexity of biological membranes, the molecular mechanisms responsible for phase separation are not easily distinguished experimentally.

The difficulty in understanding the diverging results arises, on the one hand, from the use of techniques differing in spatial (nanometers to micrometers) and temporal (nanoseconds to tens of seconds) resolution and, on the other hand, from the application of different experimental conditions. For technical reasons, domain formation was generally investigated at unsphyiological temperatures using lipids with bulky reporter groups. Both factors may affect the phase behavior of lipids. Further experiments are therefore required to test whether oganizational processes are induced by lipid domain formation under physiological conditions.

VI. INTERACTION OF MEMBRANE LIPIDS WITH AMPHIPHILIC MOLECULES AND TRANSMEMBRANE PROTEINS

A. Lipid Order Parameter in the Presence of Amphiphilic Molecules

The outer lipid membrane surface of eukaryotic cells is generally uncharged. Amphiphilic, water-soluble molecules such as local anesthetics, viral or antibiotic peptides, or peptide toxins therefore partition into the bilayer interface because of their hydrophobicity. All these compounds are found to decrease the order of lipid membranes. This is illustrated in Fig. 8 which shows the effect of incorporating the cationic peptide fragment 828–848 from the carboxy-terminus of the envelope glycoprotein gp41 of HIV-1 (P828) into bilayers composed of 1-stearoyl(d35)-2-oleoyl-*sn*-glycero-3-phosphoserine. A modest reduction of the lipid chain order near the glycerol backbone and a significant reduction towards the bilayer center are observed, indicating a decrease in the lateral packing density of the membrane and a corresponding increase of the cross-sectional area of the fatty acyl chains. An area expansion upon membrane penetration of amphiphilic compounds was also shown with molecular dynamics simulation for local anesthetics and peptides. The observation of an area increase upon insertion of local anesthetics is consistent with the phenomenon of pressure reversal of local anesthesia, which may be due to the anisotropic compression of lipid membranes under hydrostatic pressure and the consequent release of anesthetic molecules.

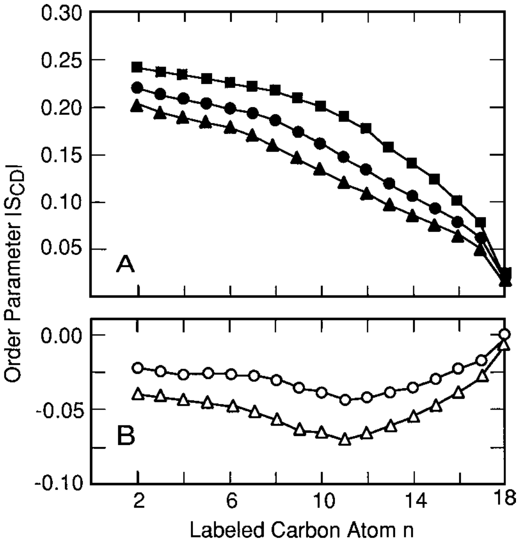


FIGURE 8 Influence of peptide P828S on the hydrocarbon chain order of 1-stearoyl$_{d35}$-2-oleoyl-*sn*-glycero-3-phosphoserine at 32°C. The smoothed order parameter profile derived from dePaked nuclear magnetic resonance powder patterns has lost the information characteristic for the beginning of the fatty acyl chains seen in Fig. 5. (A) ^{2}H NMR order parameter profiles of SOPS-d35 in the absence of P828s (■) and at lipid/peptide molar ratios of 20:1 (●) and 10:1 (▲), respectively. (B) The peptide-induced difference in order parameters along the chain at molar lipid/peptide ratios of 20:1 (○) and 10:1 (△). Peptide-induced order changes are largest in the bilayer center, suggesting that the peptide acts as a spacer that is located in the membrane's interface region. [From Smondyrev and Berkowitz (2000). *Biophys. J.* **78,** 1672.]

B. Order and Fluidity in the Presence of Transmembrane Proteins

Hydrophobic transmembrane peptides aggregate in aqueous solution and therefore do not enter a lipid membrane spontaneously. In model membranes, peptide insertion is achieved by cosolubilization of peptide and lipid in an organic solvent (detergent solution) and subsequent evaporation of the solvent (equilibrium dialysis against detergent-free buffer). Reconstitution studies show that transmembrane peptides and proteins barely perturb the lipid bilayer order, suggesting a fluid-like match between the lipid acyl chains and the outer protein surface. The investigation of hydrophobic transmembrane peptides of different lengths has led to the conclusion that the average thickness of the lipid bilayer is significantly perturbed only in cases of a large mismatch between peptide length and membrane thickness. When the hydrophobic part of the peptide was larger (smaller) than that of the pure bilayer, the membrane thickness was increased (decreased).

Larger intrinsic membrane proteins may span the membrane with several helices and perform functional tasks that can be quantified by biochemical assays. Two different approaches have been employed to study the lipid–protein interaction. One is to purify and delipidate transmembrane proteins and to reconstitute them with selectively deuterated lipids; the other is to incorporate deuterated fatty acids or other deuterated substrates into biological membranes by means of the biosynthetic pathway. In the latter case, the intact biological membrane is compared with aqueous bilayer dispersions formed from the extracted lipids. In the following we will discuss examples for the two types of assays.

Cytochrome C oxidase catalyzes the transfer of electrons from cytochrome C to molecular oxygen and is one of the best investigated intrinsic membrane proteins. The beef-heart enzyme can be purified in an almost lipid-free form and can be functionally reconstituted by incorporation into different lipid systems since the natural lipid composition is usually not required for reconstitution of an active enzyme (see Fig. 9).

The interaction of cytochrome C oxidase with lipid membranes has been investigated by means of spin-label

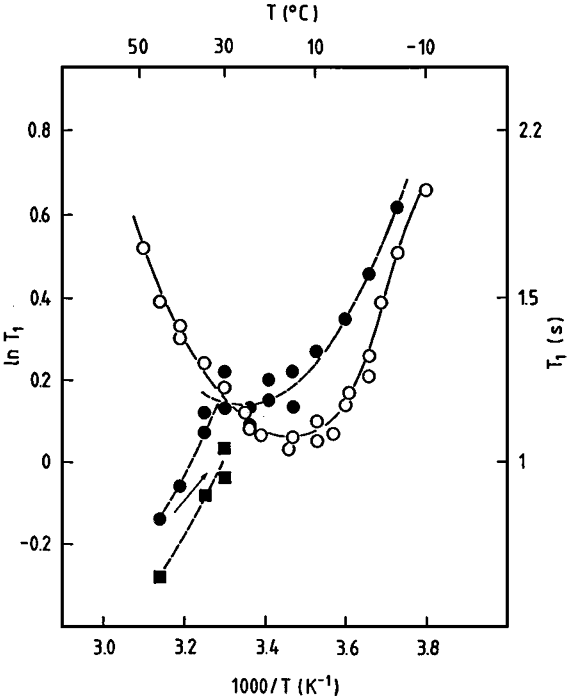


FIGURE 9 Variation of the phosphorus T_1 relaxation time with temperature. Pure POPC dispersed in 50 m*M* Tris buffer, pH 7.4, containing 1 m*M* EDTA (○). Cytochrome C oxidase functionally reconstituted with [α-$^2H_2$]POPC in 50 m*M* Tris, pH 7.4, and 10 m*M* EDTA. The higher temperatures (>30°C) were measured last (●). Same sample as in (A) but resuspended and washed in additional 10 m*M* EDTA after the measurement at 45°C (■) [From Tamm and Seelig (1983). *Biochemistry* **22,** 1474.]

electron paramagnetic resonance (epr) and by ^{2}H-, ^{14}N-, and ^{31}P-NMR experiments. The spin label method showed two motionally distinct lipid populations, with the slower component being attributed to the lipids interacting directly with the protein ("boundary lipids"). In contrast, NMR measurements of cytochrome C oxidase functionally reconstituted with headgroup and chain deuterated lipids revealed only one homogeneous population of lipids. The anisotropy of the segmental movements characterized by means of the residual ^{2}H and ^{14}N quadrupole splittings and the ^{31}P chemical shielding anisotropy as well as the segmental fluctuations, determined by measuring the ^{2}H- and ^{31}P spin-lattice (T_1) relaxation times (Fig. 10), closely resemble those of pure lipid bilayers. Taken together, the anisotropy parameters as well as the T_1 relaxation times provide no evidence for any strong polar or hydrophobic interaction between the lipid and the protein, neither in terms of a conformational change of the headgroup nor in terms of a significant immobilization of individual segments. The only noticeable difference between the NMR spectra of reconstituted membranes and pure lipid bilayers was a line broadening in the presence of protein, which probably arises from slower motions.

Similar results were obtained in reconstitution experiments with lipophilin and proteolipid apoprotein-lecithin systems, sarcoplasmic reticulum Ca^{2+}, Mg^{2+}-ATPase, rhodopsin, and glycophorin. In all these cases deuterium NMR revealed only one lipid population while the epr spectra (as far as available) showed two components. The results further show that proteins either disorder or have little effect on hydrocarbon chain order in membranes above the gel-to-liquid crystal phase transition temperature, T_c, of the pure lipids.

The question as to how an intrinsic protein affects the lipid environment was also investigated in systems containing a relatively low amount of lipid such as in partially delipidated cytochrome C oxidase surrounded by only 130 lipid molecules or in the crystalline lipovitellin/phosvitin complex containing about 100 phospholipid molecules in an interior cavity. In both systems the lipids remain in a fluid phase. Only when the lipid pool of cytochrome C oxidase was reduced to 6 to 18 molecules was a distinct broadening of the ^{2}H-NMR linewidth observed, indicating a lipid motion which was no longer axially symmetric. But even under these conditions, the total width of the spectrum was still considerably narrower than that observed for immobilized phospholipids in solid crystals.

A second, much-debated question is whether or not cardiolipins form a long-lived complex with cytochrome C oxidase. To answer this question, the remaining lipids in partially (130 lipids per protein) and highly delipidated ("lipid-depleted"; 6 to 18 lipids per protein) cytochrome C oxidase were analyzed. In the partially delipidated preparation, approximately 11 cardiolipins, 54 phosphatidylethanolamines, and 64 phosphatidylcholines were found; in the "lipid-depleted" state, the corresponding numbers are 1 or 2 cardiolipins, 3 to 8 phosphatidylethanolamines, and 2 to 8 phosphatidylcholines. This result supports a fast exchange (>10^4 s^{-1}) and is in contrast to earlier contentions that cardiolipin is the only remaining lipid in "lipid-depleted" cytochrome C oxidase. However, recent X-ray results show that the residual lipids in cytochrome C oxidase crystals are also heterogeneous and may not even contain cardiolipin. The random distribution of the remaining lipids is in accordance with a fast exchange between lipids on and off the protein surface and suggests that cardiolipin (which may have a potential role in electron transfer reactions) is at best interacting transiently rather than permanently with cytochrome C oxidase.

The results obtained in reconstitution studies were confirmed with natural membranes. The natural systems investigated are, for example, *Acholesplasma laidlawii*

(grown on a medium supplemented with specifically deuterated or perdeuterated fatty acids), cardiolipin- or glycerol-auxotroph *Escherichia coli* (grown in tissue-culture medium containing selectively deuterated fatty acids or phosphatidyl glycerol), and mouse fibroblast L-M cells (grown in tissue-culture medium containing selectively deuterated choline or ethanolamine). The membranes of these systems showed very similar fatty acid and headgroup motion, ordering, and orientation as the membranes formed from the extracted lipids without protein. No long-lived lipid–protein complexes were observed for neutral or negatively charged lipids.

VII. CONCLUDING REMARKS

Solid-state NMR measurements have shown that functional biological membranes are in the liquid crystalline state and that structural features of lipids in the crystalline phase are essentially carried over into the liquid crystalline state. An order parameter profile comparable for the most diverse membranes has been established. The absolute values of order parameters may, however, vary as much as a factor of two as a consequence of the large variation in lipid composition encountered in biological membranes. Membrane ordering *decreases* upon increasing the temperature, introducing one or several *cis*-double bonds into a saturated fatty acyl chain, or upon adding an amphiphilic guest molecule. In contrast, it *increases* up to twofold upon addition of 50% cholesterol. Transmembrane proteins barely influence the lipid order, as they perfectly match the lipid bilayer properties. Due to the action of enzymes (e.g., phospholipases) the lipid packing density and hence the membrane order may vary with time and, in turn, may modulate the function of membrane proteins. A conformational change in a membrane protein may further be induced by an out-of-plane rotation of the phospholipid headgroup dipole resulting in the development of a storage electric field across the membrane, which changes the protein structure. NMR measurements have further demonstrated that a fast exchange of lipid molecules is observed between the boundary of transmembrane proteins and the bulk lipid phase. At present, no physical–chemical evidence for the formation of domains or microdomains with lifetimes $>10^{-4}$ s has been obtained under physiological conditions.

SEE ALSO THE FOLLOWING ARTICLES

BIOENERGETICS • ELECTRON TRANSFER REACTIONS • ENERGY TRANSFER, INTRAMOLECULAR • ION KINETICS AND ENERGETICS • ION TRANSPORT ACROSS BIOLOGICAL MEMBRANES • LIPOPROTEIN/CHOLESTEROL METABOLISM • PROTEIN SYNTHESIS

BIBLIOGRAPHY

Davis, J. H. (1983). "The description of membrane lipid conformation, order and dynamics by 2H-NMR," *Biochim. Biophys. Acta* **737,** 117–71.

Devaux, P. F., and Seigneuret, M. (1985). "Specificity of lipid-protein interactions as determined by spectroscopic techniques," *Biochim. Biophys. Acta* **822,** 63–125.

Divecha, N., Clarke, J. H., Roefs, M., Halstead, J. R., and D'Santos, C. (2000). "Nuclear inositides: inconsistent consistencies," *Cell. Molec. Life Sci.* **57,** 379–393.

Koynova, R., and Caffrey, M. (1998). "Phases and phase transitions of the phosphatidylcholines," *Biochim. Biophys. Acta* **1376,** 91–145.

Mitchell, D. C., Gawrisch, K., Litman, B. J., and Salem, N., Jr. (1998). "Why is docosahexaenoic acid essential for nervous system function?" *Biochem. Soc. Trans.* **26,** 365–370.

Muniz, M., and Riezman, H. (2000). "Intracellular transport of GPI-anchored proteins," *Embo J.* **19,** 10–15.

Op den Kamp, J. A. F., Roelofsen, B., and van Deenen, L. L. M. (1985). "Structural dynamic aspects of phsophatidylcholine in the human erythrocyte membrane," *Trends Biochem. Sci.* **10,** 320–323.

Seelig, J. (1977). "Deuterium magnetic resonance: theory and application to lipid membranes," *Q. Rev. Biophys.* **10,** 353–418.

Seelig, J. (1978). "[^{2}H]Hydrogen and [^{31}P]phosphorus nuclear-magnetic-resonance and neutron-diffraction studies of membranes," *Biochem. Soc. Trans.* **6,** 40–42.

Seelig, J. (1978). "^{31}P nuclear magnetic resonance and the headgroup structure of phospholipids in membranes," *Biochim. Biophys. Acta* **515,** 105–140.

Seelig, J. (1993). "Phospholipid headgroups as sensors of electric charge," *In* "New Developments in Lipid–Protein Interactions and Receptor and Function" (K. W. A. Wirtz, ed.), Plenum Press, New York.

Seelig, J. (1995). "Metal ion interactions with lipids," *In* "Handbook of Metal–Ligand Interactions in Biological Fluids: Bioinorganic Chemistry," Marcel Dekker, New York.

Seelig, J., and Browning, J. L. (1978). "General features of phospholipid conformation in membranes," *FEBS Lett.* **92,** 41–44.

Seelig, J., and Macdonald, P. M. (1987). "Phospholipids and proteins in biological membranes: ^{2}H-NMR as a method to study structure, dynamics, and interations," *Acc. Chem. Res.* **20,** 221–228.

Seelig, J., and Seelig, A. (1980). "Lipid conformation in model membranes and biological membranes," *Q. Rev. Biophys.* **13,** 19–61.

Seelig, J., Seelig, A., and Tamm, L. (1982). "Nuclear magnetic resonance and lipid–protein interactions," *In* "Lipid-Protein Interactions" (P. Jost and O. H. Griffith, eds.), pp. 127–148, John Wiley & Sons, New York.

Natural Antioxidants In Foods

Eric A. Decker
University of Massachusetts

I. Free Radical Scavengers
II. Metal Chelators
III. Antioxidant Enzymes

GLOSSARY

Antioxidant A compounds that can inhibit oxidative processes.

Free radical A compound with an unpaired electron that can promote oxidative reaction.

Free radical scavenger A compound that can absorb a free radical to decrease the radical energy thus making it less likely to cause oxidation.

Metal chelators Compounds that can bind metals and decrease their reactivity.

Phenolic A group of chemical compounds primarily found in plants that act as antioxidant and are beneficial to health.

ATMOSPHERIC (TRIPLET) oxygen is a low energy biradical (i.e., contains two unpaired electrons). However, during metabolism of oxygen as well nitrogen, alterations can occur to produce highly reactive oxygen and nitrogen species that will react with and cause damage to biomolecules. In foods, this can cause oxidation of lipids, pigments, vitamins, and proteins, leading to off-flavor formation, discoloration, and loss of important nutrients. Foods, which are derived from a variety of different biological tissues, contain a host of different antioxidant defense systems to prevent the damaging effect of reactive oxygen and nitrogen species. However, during the processing of biological tissues into foods, the formation of oxidizing species can increase and antioxidant systems can be overwhelmed leading to uncontrolled oxidative reactions resulting in loss of quality, decrease in shelf-life, and formation of potentially toxic oxidation products. To protect food quality and safety, antioxidants are often added to processed foods. These antioxidants can be synthetically derived compounds, such as butylated hydroxytoluene and ethylene diaminetetraacetic acid. Concern over the use of synthetic food additives has driven the food industry to find effective natural antioxidants additives that are derived from biological sources. In addition, efforts to decrease oxidative deterioration have focused on the development of food processing techniques that preserve endogenous antioxidants and nutritional schemes that increase natural antioxidants in animal-derived foods.

In addition to the association of natural antioxidants with food quality, these compounds have also been associated with health benefits. The association of the protective effects of fruits and vegetables in the diet against diseases, such as cancer and cardiovascular disease, has been established for years. Comprehensive reviews on the consumption of fruits and vegetables with cancer rates have shown that 60–85% of the studies have a statistically significant association with the decrease of cancer

incidence. Individuals who consume the highest amount of fruits and vegetables have half the cancer rate as those who consume the least amount. A similar association has been seen with cardiovascular disease, with 60% of the studies reviewed showing statistically significant protective effects. The consumption of an ample supply of fruits and vegetables provides a wide variety of phytochemicals that have been shown to have health benefits and antioxidant activity. The natural antioxidants with health benefits include ascorbic acid, α-tocopherol, β-carotene, and plant phenolics.

I. FREE RADICAL SCAVENGERS

A. Phenolic Antioxidants

Phenolics are compounds that have a hydroxyl group associated with an aromatic ring structure. There are numerous variations of both natural and synthetic phenolics (see Fig. 1 for examples). Natural phenolics are found predominately in the plant kingdom. Vitamin E or α-tocopherol is a plant phenolic required in the diet of humans and other animals. Phenolic compounds primarily inhibit lipid oxidation through their ability to scavenge free radicals and convert the resulting phenolic radicals into a low-energy form that does not further promote oxidation. Chemical properties, including ability of the antioxidant to donate hydrogen to the oxidizing free radical, decrease the energy of the antioxidant radical, and prevent autoxidation of the antioxidant radical into additional free radicals, will influence the antioxidant effectiveness of a free radical scavenger (FRS). In addition, physical partitioning of phenolics will also influence their reactivity. Initially, antioxidant efficiency is dependent on the ability of the FRS to donate a hydrogen to a high energy free radical. As the oxygen–hydrogen bond energy of the FRS decreases, the transfer of the hydrogen to the free radical is more energetically favorable and thus more rapid. The ability

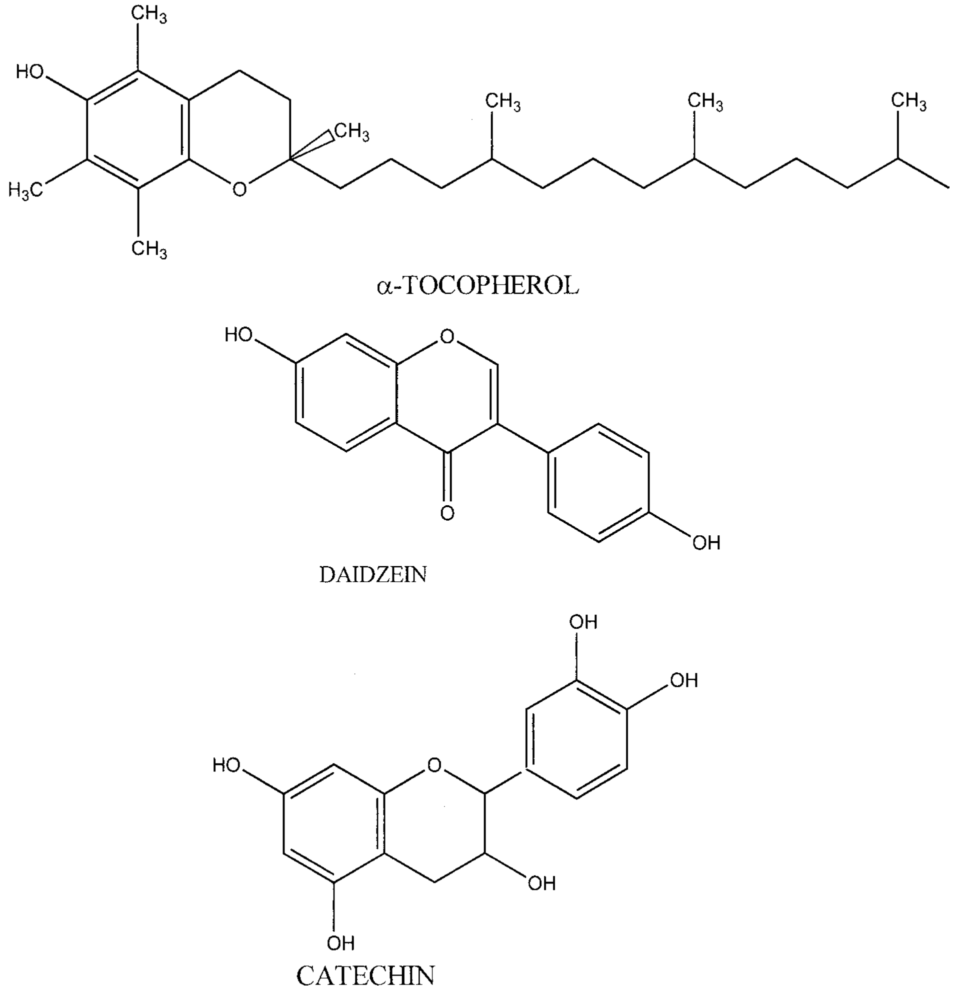


FIGURE 1 Chemical structures of some examples of phenolic antioxidants.

of a FRS to donate a hydrogen to a free radical can sometimes be predicted from standard one electron reduction potentials ($E^{\circ\prime}$). If a compound has a reduction potential lower than the reduction potential of a free radical found in a food or biological tissue (e.g., fatty acid based peroxyl radical), it can donate hydrogen to that free radical unless the reaction is kinetically unfeasible. For example, FRS including α-tocopherol ($E^{\circ\prime}$= 500 mV), urate ($E^{\circ\prime}$= 590 mV), catechol ($E^{\circ\prime}$= 530 mV), and ascorbate ($E^{\circ\prime}$= 282 mV) all have reduction potentials below peroxyl radicals ($E^{\circ\prime}$= 1000 mV, a common free radical in lipid oxidation reactions) and therefore can convert the peroxyl radical to a hydroperoxide through hydrogen donation.

The efficiency of an antioxidant FRS is also dependent on the energy of the resulting antioxidant radical. If a FRS produces a low energy radical then the likelihood of the FRS radical to promote the oxidation of other molecules is lower and the oxidation reaction rate decreases. Phenolics are effective FRS because phenolic free radicals have low energy due to delocalization of the free radical thoughout the phenolic ring structure. Standard reduction potentials can again be used to help illustrate this point. Radicals on α-tocopherol ($E^{\circ\prime}$= 500 mV) and catechol ($E^{\circ\prime}$= 530 mV) have lower reduction potentials than polyunsaturated fatty acids ($E^{\circ\prime}$= 600 mV), meaning that their radicals do not posses high enough energy to effectively promote the oxidation of unsaturated fatty acids. Effective phenolic anioxidants FRS also produce radicals that do not react rapidly with oxygen to form hydroperoxides that could autoxidize, thus depleting the system of antioxidants. Antioxidant hydroperoxides are also a problem because they can decompose into radicals that could promote oxidation. Thus, if antioxidant hydroperoxides did form, this could result in consumption of the antioxidant with no net decrease in free radicals numbers.

Antioxidant radicals may undergo additional reactions that remove radicals from the system, such as reactions with other antioxidant radicals or lipids radicals to form nonradical species. This means that each FRS is capable of inactivating at least of two free radicals, the first being inactivated when the FRS interacts with the initial oxidizing radical, and the second, when the FRS radical interacts with another radical via a termination reaction to form a nonradical product.

Phenolic compounds that act as antioxidants are widespread in the plant kingdom. Plant phenolics can be classified as simple phenolics, phenolic acids, hydroxycinnamic acid derivatives, and flavonoids. In addition to the basic hydroxylated aromatic ring structure of these compounds, plant phenolics are often associated with sugars and organic acids. The consumption of natural plant phenolics have been estimated to be up to 1 g per day. Overall, the presence of phenolics in the diet has been positively associated with the prevention of diseases such as cancer and atherosclerosis. Plant foods high in phenolics include cereals, legumes, and other seeds (e.g., sesame, oats, soybeans, and coffee); red-, purple-, and blue-colored fruits (e.g., grapes, strawberries, and plums); and the leaves of herbs and bushes (e.g., tea, rosemary, and thyme). Many natural phenolics are capable of inhibiting oxidative reactions. However, because phenolics have such a wide array of chemical structures, it is not surprising that antioxidant activities and health benefits vary greatly. Knowledge of antioxidant activity, antioxidant mechanisms, and health benefits of plant phenolics is just beginning to be understood. This section focuses on the best studied of the plant phenolics.

Tocopherols and tocotrienols are a group of phenolic FRS isomers (α, β, δ and γ; see Fig. 1 for the structure of α-tocopherol) originating in plants and eventually ending up in animal foods via the diet. Interactions between tocopherols and fatty acid peroxyl radicals lead to the formation of fatty acid hydroperoxides and several resonance structures of tocopheroxyl radicals. Tocopheroxyl radicals can interact with other compounds or with each other to form a variety of products. The types and amounts of these products are dependent on oxidation rates, radical species, lipid state (e.g., bulk vs. membrane lipids), and tocopherol concentrations.

Under condition of low oxidation rates in lipid membrane systems, tocopheroxyl radicals primarily convert to tocopherylquinone. Tocopherylquinone can form from the interaction of two tocopheroxyl radicals leading to the formation of tocopherylquinone and the regeneration of tocopherol. Tocopherylquinone can also be regenerated back to tocopherol in the presence of reducing agents (e.g., ascorbic acid). An additional reaction that can occur is the interaction of two tocopheroxyl radicals to form tocopherol dimers.

Tocopherol is found in plant foods especially those high in oil. Soybean, corn, safflower, and cottonseed oil are good sources of α-tocopherol as are whole grains (in particular wheat germ) and tree nuts. All tocopherol isomers are absorbed by humans, but α-tocopherol is preferentially transfered from the liver to lipoproteins, which in turn transports α-tocopherol to tissues. For this reason, α-tocopherol is the isomer most highly correlated with vitamin E activity.

Tea is an important source of dietary antioxidants for humans because it is one of the most common beverages in the world with annual consumption of over 40 liters/person/year. Phenolics in tea are mainly catechin derivatives, including catechin (Fig. 1), epicatechin, epicatechin gallate, gallocatechin, epigallocatechin gallate, and epigallocatechin. Tea originates from leaves harvested from the bush, *Camellia sinensis*. Processing of tea leaves

involves either blanching to produce green tea or fermenting to produce oolong or black tea. The fermentation process allows polyphenol oxidase enzymes to react with the catechins to form the condensed polyphenols that are responsible for the typical color and flavor of black teas. Green tea leaf extracts contain 38.8% phenolics on a dry weight basis with catechins contributing over 85% of the total phenolics. Condensation of catechins can decrease their solubility; therefore black tea extracts contain less phenolics (24.4%) of which 17% are catechins and 70% are condensed polyphenols (thearubigens). Extraction of phenolics with water from the leaves of rooibos (*Aspalathus linearis*) resulted in increased antioxidant activity with increasing extraction temperature and time, suggesting that brewing techniques could influence the antioxidant phenolic content of teas. Ingestion of dietary phenolics from tea has been associated with cancer prevention, and absorption of dietary tea phenolics has been reported.

Grapes and wines are also significant sources of dietary phenolic antioxidants. Grapes contain a wide variety of phenolics including anthocyanins, flavan-3-ols (catechin), flavonols (quercetin and rutin), and cinnamates (*S*-glutathionylcaftaric acid). As with many fruits, the majority of grape phenolics are found in the skin, seeds, and stems (collectively termed pomace). During extraction of juice, the pomace is left in contact with the juice for varying times in order to produce products of varying color, with increasing contact time resulting in increased phenolic extraction and, thus, darker color. Therefore, white grape juices and wines have lower phenolics contents (119 mg of gallic acid equivalents/L) than red wines (2057 mg of gallic acid equivalents/L). As would be expected, red grape juice and wines have greater antioxidant capacity due to their higher phenolic content. Both grape juice and wines have been suggested to have positive heath benefits, however, their phenolic compositions are not the same due to differences in juice preparation and changes in phenolic composition that occurs during both fermentation and storage.

The primary phenolics in soybeans are classified as isoflavones. Included among the soybean isoflavones are daidzein (Fig. 1), genistein, and glycitein, and the glycosolated counterparts daidzin, genistin, and glycitin. Unlike the phenolics in grapes and tea, soybean isoflavones are associated with proteins and, therefore, are found in soy flour and not in soybean oil. The concentration of isoflavones in soybeans varies with the environmental conditions under which the beans were grown. In addition, isoflavone concentrations in soy-based foods are altered during food processing operations such as heating and fermentation. Beside whole soybeans, isoflavones are found in soy milk, tempeh, miso, and tofu at concentrations ranging from 294–1625 μg/g product. Genistein and daidzein are absorbed into human plasma from products such as tofu and soy-based beverages. Bioavailability is low, with only 9–21% of the isoflavones being absorbed. Over 90% of the absorbed isoflavones are removed from the plasma within 24 hours.

Herbs and spices often contain high amount of phenolic compounds. For example, rosemary contains carnosic acid, carnosol, and rosmarinic acid. Crude rosemary extracts are a commercially important source of natural phenolic antioxidant additives in foods meats, bulk oils, lipid emulsions, and beverages.

B. Ascorbate

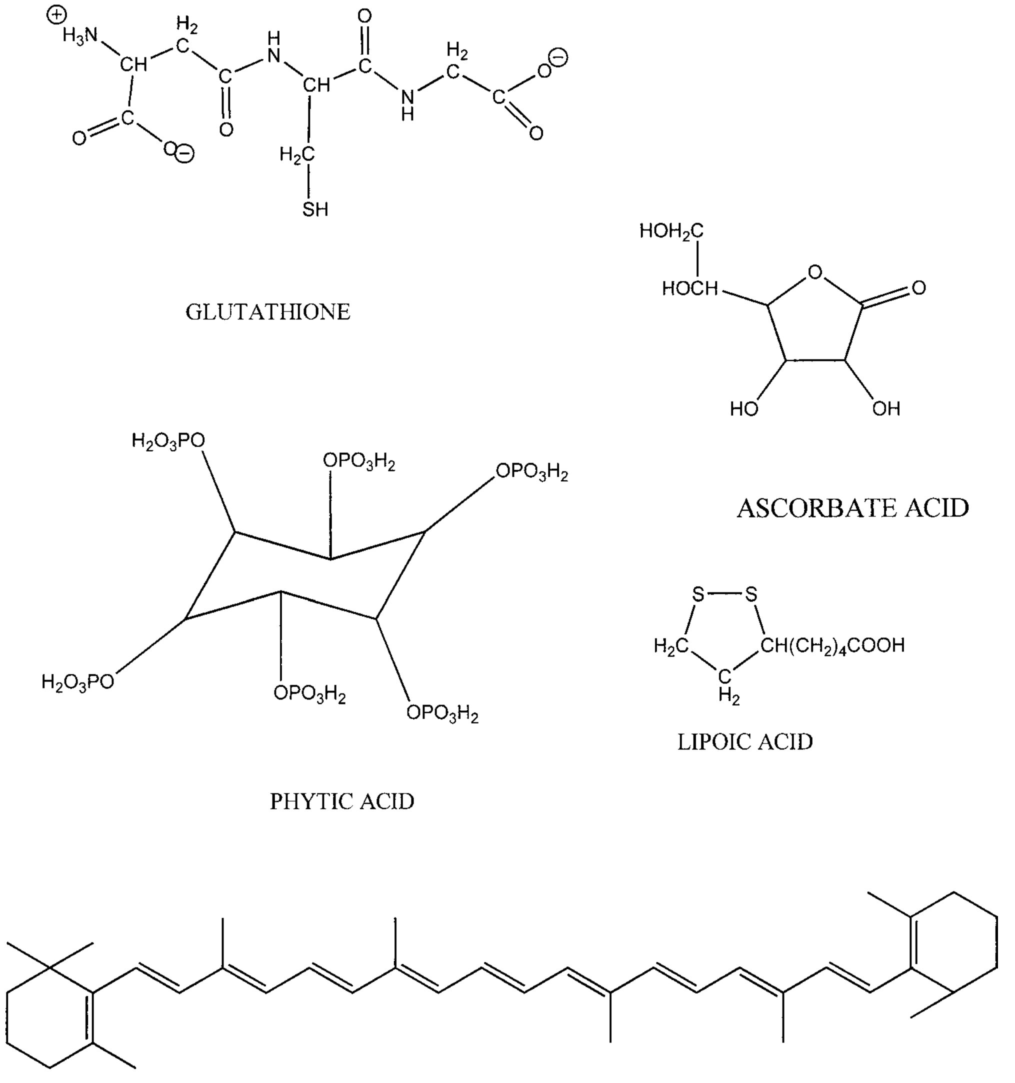
Ascorbic acid (vitamin C; Fig. 2) acts as a water-soluble free radical scavenger in both plant and animal tissues. Like phenolics, ascorbate ($E^{\circ\prime}$= 282 mV) has a reduction potential below peroxyl radicals ($E^{\circ\prime}$= 1000 mV) and thus can inactivate peroxyl radicals. In addition, ascorbate's reduction potential is lower than the α-tocopherol radical ($E^{\circ\prime}$= 500 mV), meaning that ascorbate may have an additional role in the regeneration of oxidized α-tocopherol. Interactions between ascorbate and free radicals result in the formation of numerous oxidation products. Although ascorbate seems to primarily play an antioxidant role in living tissues, this is not always true in food systems. Ascorbate is a strong reducing agent especially at low pH. When transition metals are reduced, they become very active prooxidants that can decompose hydrogen and lipid peroxides into free radicals. Ascorbate also causes the release of protein-bound iron (e.g., ferritin), thus promoting oxidation. Therefore, ascorbate can potentially exhibit prooxidative activity in the presence of free transition metals or iron-binding proteins. This does not typically occur in living tissues due to the tight control of free metals by systems that prevent metal reduction and reactivity. However, in foods the typical control of metals can be lost by processing operations that cause protein denaturation. Thus in some foods, ascorbate my act as a prooxidant and accelerate oxidative reactions.

Ascorbate is found in numerous plant foods including green vegetables, citrus fruits, tomatoes, berries, and potatoes. Ascorbate can be lost in foods due to heat processing and prolonged storage. Transition metals and exposure to air will also cause the degradation of ascorbic acid.

C. Thiols

1. Glutathione

Glutathione (Fig. 2) is a tripeptide (γ-glutamyl-cysteinyl-glycine) where cysteine can be in either the reduced or



β-CAROTENE

FIGURE 2 Chemical structures of miscellaneous natural antioxidants.

oxidized glutathione state. Reduced glutathione inhibits lipid oxidation directly by interacting with free radicals to form a relatively unstable sulfhydryl radical or by providing a source of electrons, which allows glutathione peroxidase to enzymically decompose hydrogen and lipid peroxides. Total glutathione concentrations in muscle foods range from 0.7–0.9 ug/kg. Oral administration of 3.0 of glutathione to seven healthy adults did not result in any increases in plasma glutathione or cysteine concentrations after 270 minutes. The bioavailability of glutathione in rats has also been reported to be low. Lack of, or low absorption of, glutathione may be due to the hydrolysis of the tripeptide by gastrointestinal protease.

2. Lipoic Acid

Lipoic (thioctic) acid (Fig. 2) is a thiol cofactor for many plant and animal enzymes. In biological systems, the two thiol groups of lipoic acid are found in both reduced (dihydrolipoic acid) and oxidized forms (lipoic acid). Both the oxidized and reduced forms of the molecule are capable of acting as antioxidants through their ability to quench singlet oxygen, scavenge free radicals, chelate iron, and, possibly, regenerate other antioxidants such as ascorbate and tocopherols. Lipoic and dehydrolipoic acids can protect LDL, erythrocytes, and cardiac muscle from oxidative damage.

Although lipoic acid has been found in numerous biological tissues, reports on its concentrations in foods are scarce. Lipoic acid is detectable in wheat germ (0.1 ppm) but not in wheat flour. It has been detected in bovine liver kidney and skeletal muscle. Oral administration of lipoic acid (1.65 g/kg fed) to rats for five weeks resulted in elevated levels of the thiol in liver, kidney, heart, and skin. When lipoic acid was added to diets lacking in vitamin E, symptoms typical of tocopherol deficiency were not observed suggesting that lipoic acid acts as an antioxidant in vivo. However, lipoic acid was not capable of recycling vitamin E in vivo, as determine by the fact that α-tocopherol concentrations are not elevated by dietary lipoic acid in vitamin E deficient rats.

D. Carotenoids

Carotenoids are a chemically diverse group (>600 different compounds) of yellow to red colored polyenes consisting of 3–13 conjugates double bonds and in some cases, six carbon hydroxylated ring structures at one or both ends of the molecule. ß-Carotene is the most extensively studied carotenoid antioxidant (Fig. 2). ß-Carotene will react with lipid peroxyl radicals to form a carotenoid radical. Whether this reaction is truly antioxidative seems to depend on oxygen concentrations, with high oxygen concentrations resulting in a reduction of antioxidant activity. The proposed reason for loss of antioxidant activity with increasing oxygen concentrations involves the formation of carotenoid peroxyl radicals that autoxidizes into additional free radicals. Under conditions of low oxygen tension, the carotenoid radical would be less likely to autooxidize and thus could react with other free radicals thereby forming nonradical species with in a net reduction of radical numbers.

The major antioxidant function of carotenoids in foods is not due to free radical scavenging but instead is through its ability to inactivate singlet oxygen. Singlet oxygen is an excited state of oxygen in which two electrons in the outer orbitals have opposite spin directions. Initiation of lipid oxidation by singlet oxygen is due to its electrophillic nature, which will allow it to add to the double bonds of unsaturated fatty acids leading to the formation of lipid hydroperoxides. Carotenoids can inactivate singlet oxygen by both chemical and physical quenching. Chemical quenching results in the direct addition of singlet oxygen to the carotenoid, leading to the formation of carotenoid breakdown products and loss of antioxidant activity. A more effective antioxidative mechanism of carotenoids is physical quenching. The most common energy states of singlet oxygen are 22.4 and 37.5 kcal above ground state. Carotenoids physically quench singlet oxygen by a transfer of energy from singlet oxygen to the carotenoid, resulting in an excited state of the carotenoid and ground state, triplet oxygen. Harmless transfer of energy from the excited state of the carotenoid to the surrounding medium by vibrational and rotational mechanisms then takes place. Nine or more conjugated double bonds are necessary for physical quenching, with the presence of six carbon oxygenated ring structures at the end the molecule increasing the effectiveness of singlet oxygen quenching.

In foods, light will activate chlorophyll, riboflavin, and heme-containing proteins to high energy excited states. These photoactivated molecules can promote oxidation by direct interactions with an oxidizable compounds to produce free radicals, by transferring energy to triplet oxygen to form singlet oxygen or by transfer of an electron to triplet oxygen to form the superoxide anion. Carotenoids inactivate photoactivated sensitizers by physically absorbing their energy to form the excited state of the carotenoid that then returns to the ground state by transfer of energy into the surrounding solvent.

II. METAL CHELATORS

A. Ethylene Diamine Tetraacetic Acid

Transition metals will promote oxidative reactions by hydrogen abstraction and by hydroperoxide decomposition reactions that lead to the formation of free radicals. Prooxidative metal reactivity is inhibited by chelators. Chelators that exhibit antioxidative properties inhibit metal-catalyzed reactions by one or more of the following mechanims: prevention of metal redox cycling; occupation of all metal coordination sites thus inhibiting transfer of electrons; formation of insoluble metal complexes; stearic hinderance of interactions between metals and oxidizable substrates (e.g., peroxides). The prooxidative/antioxidative properties of a chelator can often be dependent on both metal and chelator concentrations. For instance, ethylene diamine tetraacetic acid (EDTA) can be prooxidative when EDTA:iron ratios are ≤ 1 and antioxidative when EDTA:iron is ≥ 1. The prooxidant activity of some metal-chelator complexes is due to the ability of the chelator to increase metal solubility and/or increase the ease by which the metal can redox cycle.

The most common metals chelators used in foods contain multiple carboxylic acid (e.g., EDTA and citric acid) or phosphate groups (e.g., polyphosphates and phytate). Chelators are typically water soluble but many also exhibit some solubility in lipids (e.g., citric acid), thus allowing

it to inactivate metals in the lipid phase. Chelator activity is pH dependent with a pH below the pK_a of the ionizable groups resulting in protonation and loss of metal binding activity. Chelator activity is also decreased in the presence of high concentrations of other chelatable nonprooxidative metals (e.g., calcium), which will compete with the prooxidative metals for binding sites.

B. Metal-Binding Proteins

The reactivity of prooxidant metals in biological tissues are mainly controlled by proteins. Metal binding proteins in foods include transferrin (blood plasma), phosvitin (egg yolk), lactoferrin (milk), and ferritin (animal tissues). Transferrin, phosvitin, and lactoferrin are structurally similar proteins consisting of a single polypeptide chain with a molecular weight ranging from 76,000–80,000. Transferrin and lactoferrin each bind two ferric ions, whereas phosvitin has been reported to bind three. Ferritin is a multisubunit protein (molecular weight of 450,000) with the capability of chelating up to 4500 ferric ions. Transferrin, phosvitin, lactoferrin, and ferritin inhibit iron-catalyzed lipid oxidation by binding iron in its inactive ferric state and, possibly, by sterically hindering metal/peroxide interactions. Reducing agents (ascorbate, cysteine, and superoxide anion) and low pH can cause the release of iron from many of the iron-binding proteins, resulting in an acceleration of oxidative reactions. Copper reactivity is controlled by binding to serum albumin, ceruloplasmin, and the skeletal muscle dipeptide, carnosine.

C. Phytic Acid

Phytic acid or myoinositol hexaphophate is one of the primary metal chelators in seeds where it can be found at concentrations ranging from 0.8–5.3% (Fig. 2). Phytic acid is not readily digested in the human gastrointestinal tract but can be digested by dietary plant phytases and by phytases originating from enteric microorganisms. Phytate is highly phosphorylated, thus, allowing it to form strong chelates with iron, with the resulting iron chelates having lower reactivity. The antioxidant properties of phytic acid are thought to help minimize oxidation in legumes and cereal grains as well as in foods that may be susceptible to oxidation in the digestive tract. Phytic acid has been cited as a preventative agent in iron-mediated colon cancer. Although phytate may be beneficial toward colon cancer, it should be noted that it can potentially have deleterious health effects because of its ability to dramatically decrease the bioavailability of minerals including iron, zinc, and calcium.

III. ANTIOXIDANT ENZYMES

A. Superoxide Anion

Superoxide anion is produced by the addition of an electron to molecular oxygen. Superoxide anion can promote oxidative reactions by (1) reduction of transition metals to their more prooxidative state, (2) promotion of metal release from proteins, (3) through the pH dependent formation of its conjugated acid which can directly catalyze lipid oxidation, and (4) through its spontaneous dismutation into hydrogen peroxide. Due to the ability of superoxide anion to participate in oxidative reactions, the biological tissues from which foods originate will contain superoxide dismutase (SOD).

Two forms of SOD are found in eukaryotic cells, one in the cytosol and the other in the mitochondria. Cytosolic SOD contains copper and zinc in the active site. Mitochondrial SOD contains manganese. Both forms of SOD catalyze the conversion of superoxide anion (O_{2^-}) to hydrogen peroxide by the following reaction.

$$2O_{2^-} + 2H^+ \rightarrow O_2 + H_2O_2.$$

B. Catalase

Hydroperoxides are important oxidative substrates because they decompose via transition metals, irradiation, and elevated temperatures to form free radicals. Hydrogen peroxide exists in foods due to its direct addition (e.g., aseptic processing operations) and by its formation in biological tissues by mechanisms including the dismutation of superoxide by SOD and the activity of peroxisomes. Lipid hydroperoxides are naturally found in virtually all food lipids. Removal of hydrogen and lipid peroxides from biological tissues is critical to prevent oxidative damage. Therefore, almost all foods originating from biological tissues contain enzymes that decompose peroxides into compounds less susceptible to oxidation. Catalase is a heme-containing enzyme that decomposes hydrogen peroxide by the following reaction.

$$2H_2O_2 \rightarrow 2H_2O + O_2.$$

C. Ascorbate Peroxidase

Hydrogen peroxide in higher plants and algae may also be decomposed by ascorbate peroxidase. Ascorbate peroxidase inactivates hydrogen peroxide in the cytosol and chloroplasts by the following mechanism.

$$2\ \text{ascorbate} + H_2O_2 \rightarrow 2\ \text{monodehydroascorbate} + 2H_2O.$$

Two ascorbate peroxidase isozymes have been described that differ in molecular weight (57,000 versus 34,000), substrate specificity, pH optimum, and stability.

D. Glutathione Peroxidase

Many foods also contain glutathione peroxidase. Glutathione peroxidase differs from catalase in that it decomposes both lipid and hydrogen peroxides. GSH-Px is a selenium-containing enzyme that catalyzes hydrogen or lipid (LOOH) peroxide reduction using reduced glutathione (GSH):

$$H_2O_2 + 2GSH\text{—}2H_2O + GSSG$$

or,

$$LOOH + 2GSH\text{—}LOH + H_2O + GSSG,$$

where GSSG is oxidized glutathione and LOH is a fatty acid alcohol. Two types of GSH-Px exist in biological tissues, of which one shows high specificity for phospholipid hydroperoxides.

E. Antioxidant Enzymes in Foods

Antioxidant enzyme activity in foods can be altered in raw materials and finished products. Antioxidant enzymes differ in different genetic strains and at different stages of development in plant foods. Heat processing and food additives (e.g., salt and acids) can inhibit or inactivate antioxidant enzyme activity. Dietary supplementation of selenium can be used to increase the glutathione peroxidase activity of animal tissues. These factors suggests that technologies could be developed to increase natural levels of antioxidant enzymes in raw materials and/or minimize their loss of activity during food processing operations.

CONCLUSION

The biological tissues from which foods originate contain multicomponent antioxidant systems that include free radical scavengers, metal chelators, singlet oxygen quenchers, and antioxidant enzymes. Our understanding of how these endogenous antioxidants protect foods from oxidation is still in its infancy. In addition, how factors that can alter the activity of endogenous food antioxidants (e.g., heat processing, irradiation, and genetic selection of foods high in antioxidants) is still poorly understood. Finally, research is continuing to show that natural food antioxidants in the diet are very important in the modulation of disease. Thus, finding mechanisms to increase natural food antioxidants may be beneficial to both health and food quality.

SEE ALSO THE FOLLOWING ARTICLES

Food Colors • Hydrogen Bond • Lipoprotein/Cholesterol Metabolism

BIBLIOGRAPHY

Buettner, G. R. (1993). "The pecking order of free radicals and antioxidants: Lipid peroxidation, α-tocopherol, and ascorbate," *Arch. Biochem. Biophys.* **300,** 535–543.

Decker, E. A. (1998). "Strategies for manipulating the prooxidative/antioxidative balance of foods to mazimixe oxidative stability," *Trends Food Sci. Technol.* **9,** 241–248.

Decker, E. A., and Clarkson, P. (2000). "Dietary sources and bioavailability of essential and nonessential antioxidants," In: *Exercise and Oxygen Toxicity* (C.K. Sen, L. Packer, and O. Hanninen, eds.). pp. 323–358. Elsevier Science, Amsterdam.

Frankel, E. N. (1996). "Antioxidants in lipid foods and their impact on food quality," *Food Chem.* **57,** 51–55.

Graf, E., and Eaton, J. W. (1990). "Antioxidant functions of phytic acid," *Free Rad. Biol. Med.* **8,** 61–69.

Halliwell, B. (1999). "Establishing the significance and optimal intake of dietary antioxidants: The biomarker concept," *Nutr. Rev.* **57,** 104–113.

Halliwell, B., Murcia, M. A., Chirico, S., Aruoma, O. I. (1995). "Free radicals and antioxidants in foods and *in vivo*: What they do and how they work," *Crit. Rev. Food Sci. Nutr.* **35,** 7–20.

Krinsky, N. I. (1992). "Mechanism of action of biological antioxidants," *Proc. Soc. Exp. Biol. Med.* **200,** 248–254.

Liebler, D. C. (1993). "The role of metabolism in the antioxidant function of vitamin E," *Crit. Rev. Toxicol.* **23,** 147–169.

Liebler, D. C. (1992). "Antioxidant reactions of carotenoids," *Ann. N Y Acad. Sci.* **691,** 20–30.

Nawar, W. W. (1996). "Lipids," In: *Food Chemistry* (O. Fennema, ed.), 3rd edition, pp. 225–319. Marcel Dekker, New York.

Nucleic Acid Synthesis

Sankar Mitra
Tapas K. Hazra
Tadahide Izumi
University of Texas Medical Branch, Galveston

I. Structure and Function of Nucleic Acids
II. Nucleic Acid Syntheses
III. DNA Replication and Its Regulation
IV. Maintenance of Genome Integrity
V. DNA Manipulations and Their Applications
VI. Transcriptional Processes
VII. Chemical Synthesis of Nucleic Acids (Oligonucleotides)

GLOSSARY

Cell cycle Stages in the life cycle of replicating eukaryotic cells. After cell division (mitosis), a cell goes through the resting G1/Go phase prior to DNA replication in the S phase. Completion of duplication of cellular materials in the G2 phase occurs prior to mitosis (M phase).

Chromatin Cellular genome as nucleoprotein which contains DNA, histones, and a variety of nonhistone chromosomal proteins.

Chromatin remodeling Alteration in the structure of segment of chromatin which is brought about by histone acetylation/deacetylation and/or mediated by interaction with large protein complexes as a prerequisite for modulation of transcription activity.

Chromosomes Discrete and microscopically visible segments of the eukaryotic genome complexed with proteins and capped by telomeres; each normally contains thousands of genes.

***Cis* element** Short, specific DNA sequences, usually in the promoter, that bind cognate *trans*-acting factors.

Deoxyribonucleotides Monomeric units of DNA, including deoxyadenylic (dAMP), deoxyguanylic (dGMP), deoxycytidylic (dCMP), and deoxythymidylic (dTMP) acids.

DNA Deoxyribonucleic acid: linear copolymers of monomeric deoxyribonucleotides normally present as a two-stranded intertwined helix; the deoxyribose sugar moiety lacks

DNA helicase An enzyme which unwinds the double helical DNA using energy provided by ATP hydrolysis.

DNA ligase The enzyme which catalyzes joining of the $5'$ and $3'$ termini of two single-stranded DNA fragments in a double-stranded DNA by forming a phosphodiester bond.

DNA repair Enzymatic process that maintains sequence integrity by removing both endogenously and exogenously induced DNA damage. Such lesions could be

mutagenic because of misreplication at the damage site. Replication errors are also corrected by DNA repair. Repair involves removal of the DNA damage site in duplex DNA, followed by resynthesis of the damaged strand using the unaffected complementary strand as the template.

Enhancer elements DNA sequences which activate the expression of genes in an orientation- and position-independent fashion.

Episome Small extrachromosomal and sometimes self-replicating DNA molecules, including infecting viral DNA, founded in both prokaryotes and eukaryotes.

Error-bypass DNA polymerases A new class of recently discovered DNA polymerases in both prokaryotes and eukaryotes which are more tolerant of improper base pairing and may function in maintaining genomic continuity when damaged DNA bases have not been repaired.

Function The intrinsic 3′ exonuclease activity of replicative DNA polymerase or polymerase complexes needed to excise incorrect deoxynucleotides inserted at the terminus of a growing DNA chain.

Gene Basic functional unit in the genome which is transcribed to produce messenger RNA, which in turn is translated into protein. (Some genes, e.g., those for ribosomal and transfer RNAs, are only transcribed and not translated.)

Genome Complete genetic information stored in the nucleotide sequence (usually DNA) of an organism, organelle, or episome.

HMG proteins High mobility group (based on gel electrophoresis) proteins which are associated with chromatin; a subset of nonhistone chromosomal (NHC) proteins.

Lagging strand Nascent DNA strand synthesized discontinuously by replication of the $5' \rightarrow 3'$ template strand.

Leading strand Nascent DNA synthesized by continuous replication of the $3' \rightarrow 5'$ template strand.

Mitochondrial genome Multiple copies of the circular DNA duplex molecule in eukaryotic mitochondria. Believed to be a vestigial prokaryotic genome, it is replicated by a special DNA polymerase (Pol γ) which, along with other proteins required for mitochondrial DNA replication, is encoded by the nuclear genome.

Mutation Change in the genome sequence via the process of **mutagenesis**, which can occur either spontaneously due to endogenous reactions or after exposure to external mutagens, including radiation and chemicals. Mutations include large-scale sequence alterations, including deletion or insertion of thousands of DNA base pairs and genomic rearrangement which could involve **translocation** of one chromosomal segment to another. Mutations could also be subtle, including changes of a single base (known as point mutation), which include loss or addition of a single base.

Nontranscribed strand The complementary strand (5′-3′) of DNA with the same sequence as the RNA transcribed from the other (transcribed or template) strand.

Nucleosome Smallest repeat unit of chromatin nucleoprotein, containing 145 bp of DNA wrapped around a histone octamer core (2 subunits each of histone H2A, H2B, H3, and H4) along with linker DNA of variable length. Mild treatment of chromatin with DNase digests the linker and generates nucleosome fragments of different repeat lengths ("ladder").

Okazaki fragments Nascent DNA fragments generated by discontinuous synthesis of the lagging ($5' \rightarrow 3'$) strand in all organisms.

Operator A small, specific, and often palindromic DNA sequence or its repeats cognate to regulated bacterial genes. A repressor (or activator) binds the operators to prevent (or activate) transcription.

***Ori* (origin)** Origin of replication in the genome. These are unique sequences which bind the replication initiation complex as a prerequisite for primer synthesis.

PCR Polymerase chain reaction.

Plasmid Extrachromosomal DNA molecule, usually much smaller than the cell genome. Plasmids are autonomously replicated in the cell, utilizing the cellular replication machinery.

Pol DNA or RNA polymerase.

Primase Enzyme (sometimes with other accessory proteins) which is a component of the DNA replication machinery and is needed for synthesis of an oligoribonucleotide primer.

Promoter Specific DNA sequence usually found at the beginning of a gene, which binds the transcriptional machinery as a prerequisite to transcription initiation from the gene.

Replicon Unit of DNA replication in the genome, containing one *ori* site. Small genomes of bacteria, plasmids, and viruses have single replicons, while larger eukaryotic genomes have hundreds or thousands of replicons which could be simultaneously or sequentially fired for synthesis of different segments of the genome. This is necessary to reduce the overall replication time of a genome which is 10^3 times larger than a bacterial genome.

Repressor Proteins which bind to specific operators and thus negatively regulate gene expression by inhibiting transcription.

Reverse transcriptase (RT) Specialized DNA polymerase encoded by retroviruses, including the AIDS virus (HIV), which utilizes both RNA and DNA

template. It is responsible for propagation of retroviruses via synthesis of a proviral DNA intermediate.

Ribonucleotides Monomeric units of RNA, namely, adenylic (AMP), guanylic (GMP), cytidylic (CMP), and uridylic (UMP); the ribose sugar moiety of each contains a 2′-OH.

Ribosome Protein synthesis factory consisting of two differently sized subunits of ribonucleoprotein complexes with several active centers. It travels along mRNA and reads triplet codons for individual amino acids which are brought in by transfer RNAs via base pairing with cognate anticodon sequences in these RNAs. Protein synthesis occurs on the ribosome to which the growing polypeptide chain remains attached.

RNA Ribonucleic acid: linear copolymers usually of four ribonucleotides. Three major types of RNA are synthesized in the cell: ribosomal RNA (rRNA), the major component of ribosomes; transfer RNA (tRNA), the adaptor for protein synthesis; and messenger RNA (mRNA), which is required for information transfer. Other small RNAs with specialized functions are also synthesized in small amounts in both prokaryotic and eukaryotic cells.

RTPCR Reverse transcript polymerase chain reaction. Modification of the PCR method to amplify RNA, which involves generation of a complementary DNA molecule from RNA (by reverse transcriptase) which is then used in PCR.

Telomerase A special eukaryotic DNA polymerase that adds a repeat sequence to chromosome termini without a template.

Telomere Terminal region of a linear chromosome, containing partial single-stranded DNA and repeat sequences of short oligonucleotides. Its loss could cause chromosome fusion and rearrangement.

Template-independent poly(A) polymerase A template-independent RNA polymerase which catalyzes formation of AMP containing homopolymers up to several hundred monomers at the 3′ termini of nascent RNA molecules. The poly(A) tail promotes transport of mRNA from the nucleus, enhances its stability, and is necessary for translation.

Terminator Specific sequence found at the end of genes for termination of transcription due to release of RNA and RNA polymerase.

Topoisomerase Enzymes which alter topologically constrained DNA, including circular DNA, by changing the linking number. Topoisomerase I changes the linking number one at a time and does not require an external energy source. Topoisomerase II changes the linking number two at a time and generally requires ATP. The linking number is changed by transient breakage and rejoining, with an enzyme-DNA covalent bond intermediate. The enzyme acts as a swivel for rotating DNA strands around each other.

***Trans*-acting factors** Proteins that bind to specific DNA sequences (*cis* elements) in genes and regulate transcription positively or negatively.

Transcribed strand The 3′ → 5′ DNA strand utilized by RNA polymerase as its transcriptional template.

Transcriptional activator *Trans*-acting proteins which enhance transcription and, thus, the level of specific proteins.

Transcription unit Discrete segment of DNA, corresponding to one or more genes, which is utilized as a template by RNA polymerase. In prokaryotes, the transcription unit is called an operon.

Translation Synthesis of a protein, directed by mRNA molecules on ribosome.

NUCLEIC ACIDS are involved in the most fundamental processes of life. Their maintenance and production are essential in all living organisms. The hallmark of the biosphere is diversity of biological processes, even among members of the same genera, e.g., bacteria. Each organism may have some unique features in regard to nucleic acid composition, structure, and metabolism. Thus, studies on nucleic acid synthesis constitute a huge topic of research on which thousands of research articles are published each year. Therefore, it is impossible to cover all aspects of nucleic acid synthesis in this short article. Our goal is to present a broad overview of the key and general features of structure, synthesis, and processing of the various types of nucleic acids. We have limited our discussion mostly to bacteria, specifically *Escherichia coli*, and to mammals, mostly humans and mice. Most of our current knowledge has been derived from the studies of those organisms.

We have also provided appropriate references, which are mostly recent reviews. The readers should be able to peruse these for in-depth knowledge of the topics which are covered only superficially here. Finally, we have included a glossary at the beginning of this article which lists common acronyms and short descriptions of key processes and phenomena.

I. STRUCTURE AND FUNCTION OF NUCLEIC ACIDS

A. Basic Chemical Structure

The basic information for all activities in living systems, at least on our planet, is stored ultimately in nucleic acids, namely, deoxyribonucleic (DNA) and ribonucleic (RNA) acids. Except for certain viruses, DNA is the universal genetic material (Fig. 1). The chemical structures of basic

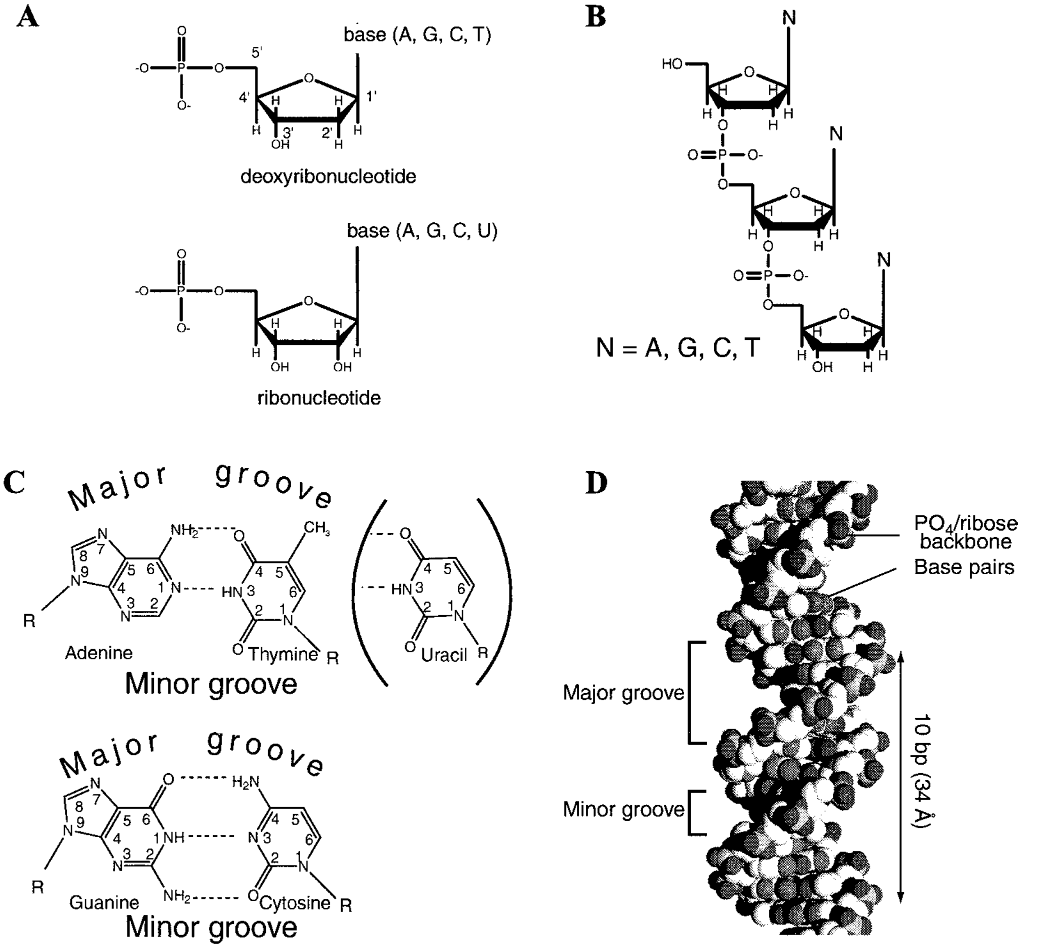


FIGURE 1 Structure of DNA and RNA: (A) structure of deoxyribonucleotides and ribonucleotides and (B) structure of polynucleotide. Each 3′ carbon of the sugar residue is linked to the 5′ carbon of the sugar residue in the next nucleotide with a phosphate to form the phosphodiester backbone. (C) Base paring of adenine with thymine (uracil) and guanine with cytosine. Dotted lines denote hydrogen bonding between two bases. R, pentose ring of nucleotide. (D) A three-dimensional structure of a DNA helix.

units of RNA and DNA have been elucidated, and both types of nucleic acids are linear polymers of monomeric units called nucleotides. A nucleotide consists of a purine or pyrimidine base linked to C′-1 of a pentose (furanose) via an *N*•*C* glycosyl bond and contains a phosphate residue attached to the sugar via an ester bond with a CH_2OH group at the 5′ position. The linear polymer in both RNA and DNA is generated by a C′-3 ester linkage of 5′ nucleotides generating a 3′-5′ phosphodiester linkage (Fig. 1B).

There are several differences in the chemical structures of DNA and RNA. First is the nature of the pentose ring in these macromolecules, i.e., ribofuranose for RNA and 2′-deoxyribofuranose for DNA (Fig. 1A). Because of the presence of deoxyribose in DNA, the monomeric unit is called a deoxyribonucleotide or simply a deoxynucleotide, while the RNA monomer unit is called a ribonucleotide. The term "nucleotide" is used generically for both RNA and DNA units. The absence of a 2′-OH group in DNA prevents alkali-mediated cleavage of the 3′-5′ phosphodiester cleavage observed in RNA and thus makes DNA more resistant to hydrolysis. Both RNA and DNA contain two types of purines, **adenine** (A) and **guanine** (G), and two types of pyrimidine bases (Fig. 1C). The second key difference between RNA and DNA is that while **cytosine** (C) is present in both RNA and DNA, RNA normally contains **uracil** (U), while DNA contains 5-methyluracil, called **thymine** (T), as the other pyrimidine base. The difference in chemical structure is reflected in the intrinsic chemical stability of these nucleic acids. The purine *N*-glycosyl bond in DNA is more unstable than in RNA, and as a result, purines are released much more easily from DNA by acid catalysis. Furthermore, cytosine deamination to produce U also occurs at a finite rate in DNA.

Various processes have evolved to maintain the genomic integrity, as discussed later.

Finally, two other critical differences between DNA and RNA are in the length and structure of the polymer chains. DNA polymers, as elaborated later, usually exist as a helix consisting of two intertwining chains, while RNA is present mostly as a single chain. Furthermore, DNA could contain up to several billion deoxynucleotide monomeric units in the genomes of higher organisms, although the genomes of smaller self-replicating units such as viruses contain only a few thousand deoxynucleotides. In contrast, RNA chains are never more than a few thousand nucleotides long.

B. Base Pairing in Nucleic Acids: Double Helical Structure of DNA

The most important discovery in molecular biology was the identification of the right-handed double helical structure of DNA, where two linear chains are held together by base pair complementarity. This discovery by Watson and Crick in 1953 heralded the era of molecular biology, which was preceded by the rapid accumulation of genetic evidence indicating that DNA, as the genetic material of all organisms, is the primary storehouse of all their information. Exceptions to this fundamental principle were found in certain bacterial, plant, and mammalian viruses, in which RNA constitutes the genome. However, the viruses are obligate parasites and are not able to self-propagate as independent species; thus, they have to depend on their hosts, which have DNA as their genetic material. Thus, DNA in all genomes (except some single-stranded DNA viruses) consists of two strands of polydeoxynucleotides which are anti-parallel in respect to the orientation of the 5′-3′ phosphodiester bond in the polymers (Fig. 1D). The two strands are held together by H-bonding between a purine in one strand and a pyrimidine in the complementary strand. Normally, adenine (A) pairs with T and G pairs with C; A and T are held together by two H-bonds, and G and C are held together by three H-bonds involving both exocyclic C═O and ring NH (Fig. 1C). As a result, G•C pairs are more stable than A•T pairs. Because U is structurally nearly identical to T, except for the C-5 methyl group, U also pairs with A in the common configuration. Although H-bonds are inherently weak, the stacking of bases in two polynucleotide chains makes the duplex structure of DNA quite stable and induces a fibrillar nature in the DNA polymer. X-ray diffraction studies of the DNA fiber, and subsequent crystallographic studies of small (oligonucleotide) DNA pieces, led to the detailed structural elucidation. This was initially aided by chemical analysis showing equivalence of purines and pyrimidines in all double-stranded DNA and equimolar amounts of A and T and of G and C (Chargaff's rule), unlike in RNA, which is single stranded (except in some viruses). X-ray diffraction studies also showed that DNA in double helix exists in the B-form, which is right handed and has a wide major groove and a narrow minor groove. Most of the reactive sites in the bases, including C═O and NH groups, are exposed in the major groove (Figs. 1C and 1D). One turn of the helix has10 base pairs (bp) with a rise of 34°. Thus, each pair is rotated 36° relative to its neighbor. Elucidation of the structure of DNA bound to proteins show that one turn of the helix containing 10.5 bp could be significantly bent or distorted. For example, some DNA binding proteins bind to the minor groove, causing its widening accompanied by compression of the major groove. In some special regions of the genomes, e.g., in telomeres and segments with unusual repeated sequences, alternative forms such as triple helical structure and Z-DNA may exist. The Z-DNA has a left-handed, double-helical structure. In these or in torsionally stressed DNA, the bases can be held together by different type of H-bonding called Hoogsteen base pairing.

C. Size, Structure, Organization, and Complexity of Genomes

Except for certain viruses, DNA is the genetic material for all organisms and self-replicating units, including viruses and such intracellular organelles as chloroplasts (in plants), kinetoplasts (in protozoa), and mitochondria (in most eukaryotes). Genomic DNA is double helical (except for the genomes of certain bacterial viruses), and its size is related to the complexity of the organism (Table I). In subcellular organelles, viruses, and plasmids, the genome often exists as a circular molecule consisting of up to several thousand base pairs. The genome of bacteria, such as that of the widely studied enteric strain *E. coli*, is present as a single, circular, double-stranded molecule containing about 4.7 million base pairs. By and large, the genome of many small self-replicating entities is circular DNA, without any terminus in the unbranched polymeric chain.

In contrast, the large nuclear genomes of more complex organisms (from lower eukaryotes such as unicellular yeast with a genome size only an order of magnitude larger than that of *E. coli*, to mammals with genomes larger by three orders of magnitude) consist of multiple, distinct, linear subunits organized in chromosomes. Depending on the stage of the cell cycle, the structure of **chromosomes** (collectively called **chromatin**) varies from the highly extended and amorphous state occurring in much of the (**interphase**) nucleus to highly compacted, linear, organized chromosomes (**metaphase**) after completion of DNA duplication followed by cell division (**mitosis**). This

TABLE I Genomic DNA Characterized in Biology[a]

Organism	Structure	Total size (bp)	Number of genes	Sequence
Bacteriophage	Linear, circular	$5 \sim 200 \times 10^3$	10∼100	Completed for many species
Virus		Up to 2×10^5	10∼100	Completed for many species
Bacteria *E. coli*	Circular	4.6×10^6	∼4300	Completed
Eukaryote				
yeast (*S. cerevisiae*)	Linear	1.4×10^7	∼6000	Completed
Drosophila	Linear	1.4×10^8	1.4×10^4	Partially completed
Human	Linear	3×10^9	4×10^4 to 1×10^5	Partially completed

[a] As of Feb 2001 the data are to be renewed continuously and are available at the website http://ncbi.nlm.nih.org/entrez.

complex organization of eukaryotic genomes is a distinctive feature which separates them from the prokaryotes.

D. Information Storage, Processing, and Transfer

The central dogma of molecular biology is that information is transferred from DNA to RNA to proteins. The proteins (which include the enzymes and structural components of cells) are directly responsible for most cellular activities and functions. The information needed for all functions of all organisms is stored in the genomic DNA sequence, which contains discrete units defined as **genes**. Each gene encodes a protein whose function and activity are determined by its primary sequence. The discovery of colinearity of the DNA nucleotide sequence and the amino acid sequence of the encoded polypeptide in prokaryotes and their viruses led to the discovery of the **genetic code** which postulates that a three-nucleotide sequence in DNA, called a **codon**, is responsible for insertion of a specific amino acid in the polypeptide chain during its synthesis.

Thus, the information content in the genomic DNA of a cell needs not only to be preserved and passed on to the progeny cells during replication, an essential characteristic and requirement of all living organisms, but also has to be processed and transferred via proteins to the ultimate cellular activities, including the metabolism.

Elucidation of the double-helical structure of DNA lends itself to an elegant but simple mechanism of perpetuation of the DNA information during duplication, called semi-conservative replication. In this model (Fig. 2), the two strands of DNA separate, and each then acts as the template for synthesis of a new daughter strand based on base pair complementarity and strand polarity. Thus, the two strands of the DNA double helix, though not identical in sequence, are equivalent in information content.

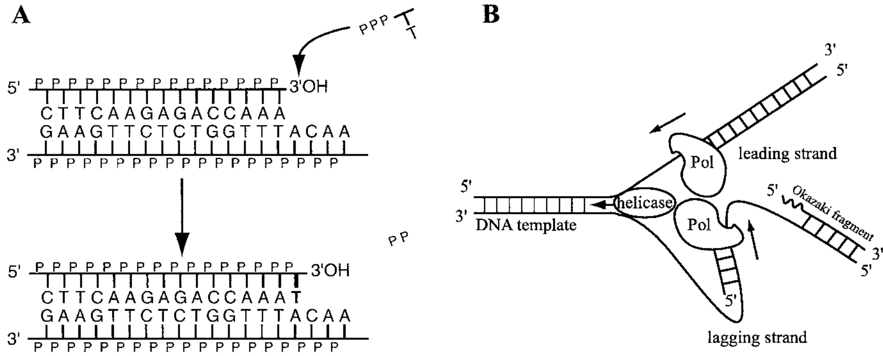


FIGURE 2 DNA polymerization reaction. (A) According to the base pairing rules, a deoxythymidinetriphosphate (dTTP) is added at the 3′-OH end of the top strand through a transesterification reaction catalyzed by a DNA polymerase. (B) Two units of DNA polymerase form a heterodimer complex to carry out replication in a semi-conservative way. Because the reaction goes only in the 5′ → 3′ direction, one side (the leading strand) is synthesized continuously, while the other (the lagging strand) consists of short DNA fragments (Okazaki fragment). DNA replication is initiated by an RNA primer (waved line) which is synthesized by a primase. There are a number of accessory but essential proteins besides the polymerase unit.

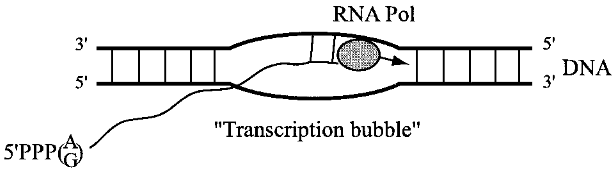


FIGURE 3 An RNA polymerase unit (filled circle), which consists of multiple factors, opens DNA helix (shown as a bubble) and synthesizes RNA in the $5' \rightarrow 3'$ direction.

The intermediate carrier in the transfer of information from DNA to protein is the **messenger RNA** (mRNA), which is **copied** (transcribed) from only one of the two strands (Fig. 3), based on base pair complementarity (except for the presence of U in RNA in the place of T; Fig. 1C). In the synthesis of both DNA (**replication**) and RNA (**transcription**), the polynucleotide chains are synthesized by sequential addition of monomeric units (deoxyribonucleotide for DNA and ribonucleotides for RNA) to the 3′ end of the growing chain (Fig. 3).

The mRNA is read out by ribosomes, the ribonucleoprotein complex which functions as the factory for protein synthesis. The codons are recognized as blocks because they code for specific amino acids. Thus, the linear polypeptide sequence is determined by the linear mRNA sequence.

E. Chromosomal DNA Compaction and Its Implications in Replication and Transcription

Metaphase chromosomes in cells undergoing mitosis are visible under the light microscope. Their formation requires some 10^4- to 10^5-fold condensation of uninterrupted linear duplex DNA which has a 2-nm diameter. Such **compaction** is accomplished in a highly complex and stepwise fashion. Because DNA is a polyelectrolyte with two negative charges per nucleotide, charge neutralization and shielding is required before the polymer can be folded in an ordered, condensed structure. In addition to metal ions and polyamines, the major source of the positive charge in chromatin is the family of highly basic small proteins, called histones, which are rich in the basic amino acid residues lysine and arginine needed to neutralize the charge of the phosphate backbone of DNA. The prokaryotes also have basic proteins (such as HU protein in *E. coli*) which induce DNA condensation. However, **chromatin** compaction in eukaryotes is carried out in stages. The simplest folded unit of DNA is the 10-nm nucleosome, consisting of a core histone octamer containing two molecules each of histone H2A, H2B, H3, and H4 around which nearly two turns of the DNA is wrapped. The nucleosome cores are connected by a stretch of linear DNA (linker) of variable length which is covered by histone H1 or H5. The polymeric chain nucleosomes are then folded in a 30-nm fiber whose structure is stabilized by the interaction among histones and a number of other proteins collectively called **nonhistone chromosomal proteins** (NHC), including **high mobility group** (HMG), which are not particularly basic. Eventually, the fibers are condensed into highly compacted metaphase chromosomes. The nature of the interactions present in interphase and metaphase chromosomes is not clear.

However, the implications of this compaction are profound. It is absolutely essential to condense the mammalian genome, which in an extended linear form more than 1 m long, to a volume which can be accommodated in the nuclear volume of 10–30 femtoliters. At the same time, the genes will be buried in condensed chromatin, and yet their specific sequences need to be exposed for various processes of information transfer. Thus, for both transcription and replication, the chromatin has to be decondensed. This was evident in early *in vitro* studies which showed that both these processes are severely inhibited when DNA is complexed with histones.

F. DNA Sequence and Chromosome Organization

The massive human genome project should achieve its goal of determining the complete sequence of human and mouse genomes in the near future; a "rough draft" has already been obtained. Furthermore, this genome initiative, pursued by both government and private enterprises in the United States and other countries, has already culminated in elucidating the complete sequence of *E. coli* and other bacteria, as well as yeast, a nematode, and the fruitfly *Drosophila melanogaste*. Significant progress has been made in elucidating the nucleotide sequences of both human and mouse genomes by using a two-pronged approach. On one hand, the sequences of transcribed regions of the genomes are being deduced from sequences of randomly isolated mRNA segments reverse transcribed into DNAs. At the same time, complete DNA sequences of fragments of whole chromosomes are being directly determined. This has opened up a huge scientific challenge of deciphering the genetic information, identifying unknown genes and their encoded proteins, and the variability of gene sequences with corresponding changes in the protein sequences in individuals. Functional genomics is a newly created discipline which deals with the deterministic prediction of protein functions from the primary sequences. One extension of such analysis is to ascertain the consequences of allelic polymorphisms in the human genome, i.e., minor changes in the sequences of cellular proteins which do not cause an explicit pathological phenotype and yet

may affect survival and predisposition to specific diseases in the long term.

G. Repetitive Sequences: Selfish DNA

Even before the precise genome sequences are elucidated, one unique feature of the metazoan DNA sequence has been established from a number of studies. A large fraction (perhaps up to 90% or more) of the total genomic sequence in metazoan cells do not encode any information. Some of these sequences are present as noncoding intervening regions in genes, named “introns,” which do not code for proteins. However, the intron sequences are transcribed but are removed during processing (“splicing”) to generate mature mRNA, as discussed later. Many of the other genomic sequences are not even transcribed, and these may often be present as multimeric repeats of shorter units. These repetitive sequences have no known function in the cell, yet are maintained and replicated as an integrated part of the genome; such DNA is referred to as “selfish DNA.”

Metaphase chromosomes are organized in substructures distinguished by their staining with dyes. **Euchromatin** regions contain transcribed sequences, while **heterochromatin** regions contain large segments of repetitive sequences. Metaphase chromosomes are also characterized by specific stained sequences (named **centromeres**) in the middle of the elongated structure, in addition to telomeres at the termini, as discussed earlier. Both centromeres and telomeres have unique repetitive sequences, and in some cases similar sequences have been observed in other regions of chromosomes; these regions are highly condensed and not transcribed.

H. Chromatin Remodeling and Histone Acetylation

In order to make the DNA template available for both replication and transcription, the chromatin is “remodeled.” One way to accomplish this reversible process is by altering the electrostatic interaction with histone. Acetylation of lysine residues (and to some extent phosphorylation of serine and threonine residues) reduces the binding affinity of histones with DNA in nucleosome cores and may thus allow exposure of free DNA to the transcriptional machinery. Additionally, a more complex energy-driven process involving the proteins SNF1 and SWI causes a major alteration of the chromatin structure, which is necessary for reprogramming of the transcriptional regimen during growth, development, and associated differentiation. DNA replication also requires access of DNA in free form to the replication machinery and, therefore, may also be dependent on the same remodeling process and could even require dissociation and reassociation of the nucleosome core.

II. NUCLEIC ACID SYNTHESES

A. Similarity of DNA and RNA Synthesis

All nucleic acids are usually synthesized by DNA template-guided polymerization of nucleotides—ribonucleotides for RNA and deoxy(ribo)nucleotides for DNA. The reactant monomers are 5′ ribonucleoside (or deoxyribonucleoside) triphosphates. These can be described in the following chemical equations:

$$\text{DNA} + \text{ndNTP} \rightleftarrows \text{DNA} + \text{nPP}_\text{i}$$

and

$$(\text{DNA}) + \text{nNTP} \rightleftarrows (\text{DNA}) + \text{RNA} + \text{nPP}_\text{i}.$$

Enzymatic polymerization is carried out by DNA and RNA polymerases, both of which carry out pyrophosphorolysis, i.e., cleavage of a high energy pyrophosphate bond coupled to esterification of 5′ phosphate linked to the 3′-OH of the previous residue. The reaction is reversible, although it strongly favors synthesis. Degradation of nucleic acids is not due to reversal of the reaction, but rather a hydrolytic reaction catalyzed by nucleases, namely, RNases and DNases, which generate nucleotides or deoxynucleotides, respectively.

Three distinct stages are involved in the biosynthesis of both DNA and RNA: **initiation**, **chain elongation**, and **termination**. Initiation denotes *de novo* synthesis of a nucleic acid polymer which is generally well regulated by complex processes, as described later. The key difference in initiation of a DNA vs RNA chain is that RNA polymerases can start a new chain, while all DNA polymerases require a “primer,” usually a short RNA or DNA sequence with a 3′-OH terminus, to which the first deoxynucleotide residue is added. Elongation denotes continuing polymerization of the monomeric nucleotides, and termination defines stoppage of nucleotide addition to the growing polymer chain.

During synthesis the enzymes catalyzing the polymerization reaction are guided by nucleic acid templates that provide the complementary sequence for the incorporated nucleotides (Fig. 4). The basic catalytic enzyme in such reactions is called DNA or RNA polymerase. In cells the template for both DNA and RNA is genomic DNA. There are some exceptions to these general rules. Some DNA polymerases can synthesize homo- or heteropolymers of deoxynucleotides *in vitro* in the absence of a template; the substrate is restricted to one or two dNTPs. While it is unlikely that these homo- or heteropolymers, e.g.,

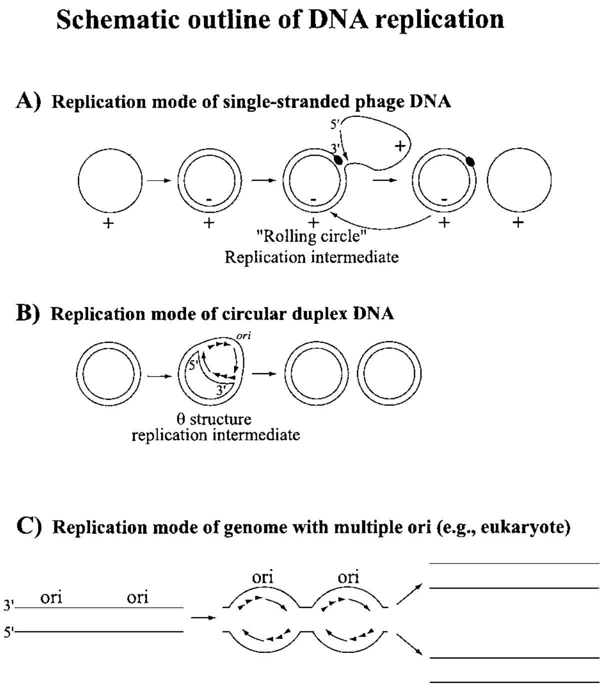


FIGURE 4 Replication of circular DNA of prokaryotes and viruses, plasmids, and mitochondria. The basic steps of replication are shown. (**A**) Rolling circle mode of replication for single-stranded circular DNA: single-stranded (ss) DNA is replicated to the replicative form (RF), which then acts as the template for progeny ssDNA synthesis via a rolling circle intermediate. (**B**) Circular duplex DNA can be replicated at the *ori* site by formation of a θ intermediate. Replication could be bidirectional (as shown here) or unidirectional. $5' \rightarrow 3'$ chain growth dictates that DNA synthesis is continuous on one side of the *ori* and discontinuous on the other side for each strand; (+) and (−) strands are shown to distinguish the strand types. (**C**) Replication of a linear genome with multiple origins.

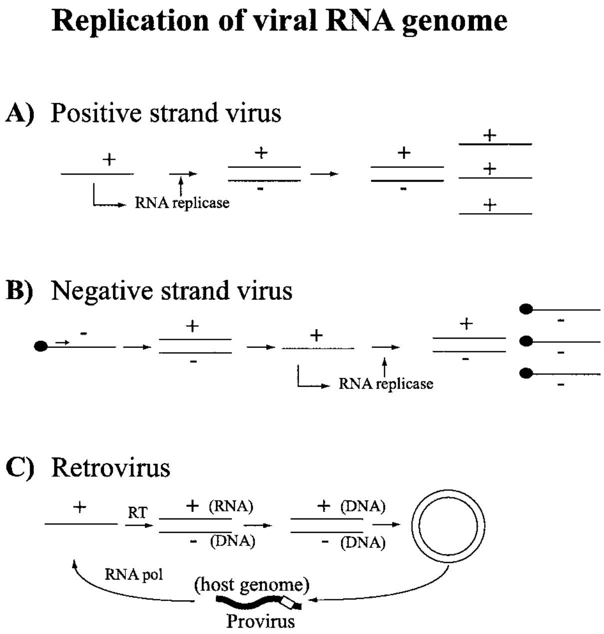


FIGURE 5 Replication of mammalian viral RNA genome. The basic steps of replication are shown for (**A**) a (+) strand genome, which acts as an mRNA for encoding viral proteins; (**B**) a (−) viral genome cannot encode protein and first has to be replicated by the RNA replicase (•) which is present in the virus particle. Once the complementary (+) strand which serves as mRNA is synthesized, viral-specific proteins are synthesized, including RNA replicase. (**C**) Replication of (+) stranded retroviral genomes first involves synthesis of the reverse transcriptase which directs synthesis of duplex DNA in two stages from the RNA template. Circularization of the DNA followed by its genomic integration allows synthesis of progeny viral RNA by the host transcription machinery.

$(dA\bullet dT)_n$ or $poly(dA)_n\bullet poly(dT)_n$, are formed *in vivo*, the availability of these polymers significantly advanced our understanding of the properties of DNA, before the age of chemical or enzymatic oligonucleotide synthesis.

There are some exceptions to the norm of DNA-dependent DNA or RNA synthesis, mostly in lower eukaryotes or viruses (Fig. 5). One example is RNA-dependent RNA synthesis in plant, animal, or bacterial viruses. In these cases, a single-stranded RNA template rather than double-stranded DNA guides synthesis of the complementary RNA strand, based on conventional base pairing. The polarity of RNA adds a level of complexity during synthesis. Thus, the RNA genome of a virus that can be directly read and thus provides the mRNA function is called the positive strand, as in polio virus. In this case, the viral genome RNA functions as the mRNA and encodes the **RNA polymerase**, which is synthesized like other viral proteins in the infected cell. This RNA polymerase subsequently synthesizes the complementary negative strand, which then serves as the template for synthesis of the progeny positive strand RNA. The progeny RNA is then packaged into mature progeny virus.

In contrast, the genomic RNA of **negative strand viruses** (e.g., vesicular stomatitis virus) cannot function directly as mRNA and thus cannot guide synthesis of proteins, including the **RNA replicase**, by itself after the infection of host cells. These viruses carry their own RNA replicase within the virion capsids, which carry out (+) mRNA strand synthesis after infection (Fig. 5).

Retroviruses comprise diverse groups of viruses, including human immunodeficiency virus (HIV), which share a common mechanism of genome replication. The RNA genomes of these viruses encode an **RNA-dependent DNA polymerase** (reverse transcriptase or RT) which first generates the **complementary (c) DNA** of the viral genome. RT has also RNaseH (specific nuclease for degrading RNA from RNA–DNA hybrids) and DNA-dependent DNA polymerase activities. After copying the RNA template, the enzyme degrades the RNA and is able to convert the resulting single-stranded cDNA to duplex

DNA. This is then integrated into the host cell genome as proviral DNA, from which the progeny viral RNA is eventually transcribed. Thus, the reverse transcriptase is an unusual polymerase because it can utilize both RNA and DNA templates (Fig. 5). There is strong evidence that such reverse transcription was involved in synthesis of "retrotransposons," a special class of mobile genetic elements, during the evolution of mammalian genomes. These mobile genetic elements, also known as transposons, when identified in bacteria and lower eukaryotes, consist of specific DNA sequences which can be relocated randomly in the genome. The transposition is mediated by enzymes called transposase, usually synthesized by a gene in the transposon. During transposition of retransposons, certain mRNAs are reverse transcribed and then integrated into the genome like the proviral sequence. The presence of specific flanking sequences allows these elements to relocate to other sites in the genome.

B. DNA Replication vs Transcription: Enzymatic Processes

The broad chemical steps in DNA and RNA synthesis are quite similar, in that both processes represent reading of a DNA strand as the template. However, while both strands of DNA have to be copied, transcription is polar because only one strand is normally copied into RNA whose sequence is identical to the other strand (except for replacement of thymidine by uridine). This is achieved by the presence of discrete start and stop signals bracketing "**transcription units**" corresponding to each gene containing unique sequences, called **promoters**; their sequences provide the recognition motif for RNA polymerase to bind and start RNA synthesis unidirectionally. Similarly, the stop sequences are recognition motifs for the transcription machinery to stop and fall off the DNA template.

As mentioned before, the two strands of a DNA double helix are of opposite polarity, i.e., one strand is in the $5' \rightarrow 3'$ orientation and its complementary strand in the $3' \rightarrow 5'$ orientation. Furthermore, the fact that all nucleic acid polymerases can polymerize nucleotide monomers only in the $5' \rightarrow 3'$ direction as guided by base pairing with a template does not pose a problem for RNA synthesis because only the $3' \rightarrow 5'$ strand of the DNA template is copied. However, DNA replication, where both strands have to be copied in the same $5' \rightarrow 3'$ direction of the duplex template, introduces a complication situation (Figs. 2 and 5). The $3' \rightarrow 5'$ strand is copied like RNA, while the $5' \rightarrow 3'$ strand has to be copied in the opposite direction. It has been observed in all cases that simultaneous replication of both strands is accomplished by continuous copying of the $3' \rightarrow 5'$ strand, also called the **leading strand**, while the $5' \rightarrow 3'$ strand is copied after a brief delay when separation of the strands occur, so this nascent strand is called the **lagging strand** (Fig. 2). The leading strand can be synthesized continuously without interruption, while the lagging strand is synthesized discontinuously after the leading strand is synthesized. The discontinuous fragments are also called **Okazaki fragments**, named after its discoverer.

C. Multiplicity of DNA and RNA Polymerases

Multiple DNA and RNA polymerases are present in both eukaryotes and prokaryotes, which evolved to fulfill distinct roles in the cell. In *E. coli*, DNA polymerases I (Pol I), II (Pol II), and III (Pol III) account for most DNA polymerase activity. Pol I has the highest enzymatic activity and was the first DNA polymerase to be discovered by A Kornberg. However, Pol III is responsible for cellular DNA replication, while Pol I is involved in gap filling necessary during normal DNA replication (to fill in the space of degraded RNA primers) and also during repair of DNA damage. Pol II and two other DNA polymerases, Din B and UmuD/C, are responsible for replication of damaged DNA when it remains unrepaired.

Eukaryotic cells express five different DNA polymerases, α, β, γ, δ, and ε, for normal DNA replication and repair. Pol α is involved in synthesis of primers for DNA replication; Pol β and possibly Pol ε are involved in repair replication of damaged DNA. Pol δ (and possibly Pol ε) are responsible for replication of the nuclear genome. Pol γ found in the mitochondria is responsible for replication of the mitochondrial genome. Several additional DNA polymerases recently identified and characterized are involved in replication of unrepaired damaged bases, like the *E. coli* DinB and UmuD/C (Table II).

E. coli has only one RNA polymerase, while eukaryotes have three distinct RNA polymerases, Pol I, Pol II, and Pol III, which transcribe different types of genes. RNA Pol I makes only ribosomal RNAs, which constitute the largest fraction of total RNA and, in fact, a significant fraction of the cellular mass. Pol III transcribes small RNAs, including transfer RNAs, which function as carriers of cognate amino acids and are required for protein synthesis. RNA Pol II transcribes all genes to generate mRNA, which encodes all proteins. Thus, this enzyme recognizes the most diverse group of genes. All of these RNA classes are initially synthesized as longer precursors that require extensive, often regulated, processing to yield the mature RNA product.

RNA and DNA polymerases encoded by virus and other episomal genomes are, in general, smaller and have fewer subunits than the cellular polymerases. Cellular polymerase holoenzymes are rather complex with multiple

TABLE II Cellular DNA Polymerases

Prokaryote (*E. coli*)	***In vivo* function**
Pol I	Nonreplicative removal of 5′ primer of Okazaki fragments
Pol II	Nonreplicative, damage responsive polymerase
Pol III	Replicative synthesis
Din B	Lesion bypass DNA synthesis
UmuC	Lesion bypass DNA synthesis
Eukaryote	
Pol α	RNA primer synthesis
Pol β	Repair synthesis
Pol δ	Replicative (repair) synthesis
Pol ε	Replicative (repair) synthesis
Pol ζ	Damage bypass synthesis
Pol η	Damage bypass synthesis
Pol θ	Damage bypass synthesis
Pol ι	Damage bypass synthesis
Pol γ	Mitochondrial DNA synthesis

subunits which may have distinct functions. These will be discussed later.

III. DNA REPLICATION AND ITS REGULATION

A. DNA Replication

DNA replication is initiated at discrete sequences called **origin** (*ori*) of replication to which DNA polymerase and accessory proteins bind and copy both strands, as predicted by the semi-conservative replication model (Fig. 2B). In contrast to unidirectional RNA synthesis, DNA replication in most genomes occurs bidirectionally (Fig. 2B). This results in both continuous and discontinuous synthesis of the same strand on two sides of the origin of replication. Some circular genomes, such as mitochondrial DNA, are replicated unidirectionally. In these cases, replication starting at the *ori* proceeds **continuously** in the $5' \rightarrow 3'$ direction, followed by **discontinuous** synthesis of the complementary strand. Termination occurs at the same site as the *ori* after the circle is completely traversed. During replication of the mitochondrial genome, elongation of the continuous strand pauses at some distance from the *ori*, resulting in a **bubble** (θ structure) structure named a **D-(displacement) loop** (Fig. 4A).

The single-stranded DNA genomes of certain small *E. coli* viruses (such as M13 and ϕX174) are replicated in the form of rolling circles in which unidirectional synthesis of one (virus genome) strand occurs by continuous displacement from the template (complementary strand; Fig. 4A). The initial duplex DNA (called the replicative form or RF) is the template for **rolling circle synthesis** and is formed first by replication of the single-stranded form. Such a single-stranded circular DNA template has been exploited in recombinant DNA techniques.

Small organisms (e.g., bacteria), as well as plasmids and many viruses, have only one *ori* sequence per cellular genome (4.7×10^6 nucleotide pairs in *E. coli*), which is often an uninterrupted DNA molecule (Figs. 4A and 4B). In complex organisms, with a much larger genome size ($\sim 3 \times 10^9$ nucleotide pairs for mammals), which is divided into multiple discrete chromosomes, thousands of *ori* sequence are present (Fig. 4C), although not all of them may be active in all cells; this requires that replication be regulated and coordinated.

B. Regulation of DNA Replication

Semi-conservative replication of the genome ensures that each daughter cell receives a full complement of the genome prior to cell division. In eukaryotes, this is achieved by the distinct phases of the cell cycle, namely, G1 phase, during which cells prepare for DNA synthesis; S phase, in which DNA replication is carried out; and G2-M (mitosis), during which the replicated chromosomes segregate into the two newly divided daughter cells. Unlike in eukaryotes, DNA replication in prokaryotes may occur continuously during growth (in rich medium). Thus, the copy number of genomes could exceed two in rapidly growing cells. In the case of viruses, which multiply by utilizing the host cell synthetic machinery and eventually killing them, genome replication may be not controlled. However, plasmid DNA, as well as the genomes of organelles such as mitochondria and chloroplasts, is replicated with some degree of regulation. In these cases the genomic copy number can vary within limits as a function of growth condition.

C. Regulation of Bacterial DNA Replication at the Level of Initiation

In all organisms, as well as autonomously replicating DNA molecules of organelles and plasmids, replication is divided into three stages: initiation, chain elongation, and termination. The control of replication occurs primarily at the level of initiation of DNA synthesis at the "origin" (*ori* site). Because DNA chains cannot be started *de novo* and requires a primer, the initiation complex contains primase activity for synthesis of an RNA primer. Discontinuous synthesis of Okazaki fragments needs repeated primer synthesis for each fragment as an integral component of chain elongation. Initiation of the primer at the *ori* sequence rather than elongation of initiated chains is the critical event in DNA replication control.

Different replicons of prokaryotes and eukaryotes utilize distinct mechanisms which vary in complexity, depending on the complexity of the organisms. A common feature of replication initiation control in *E. coli* genomes and plasmids is the presence of repeats of A•T rich sequences which facilitate unwinding of DNA and one or multiple repeats of a "dnaA box" to which the initiator **DnaA** protein in *E. coli* or its functional homolog (called **Rep** in other cases) binds to allow helical unwinding and primer synthesis. The level of DnaA protein regulates the initiation frequency and, in turn, is controlled at the level of transcription of the *dnaA* gene. Thus, there are complex negative autofeedback loops to control dnaA gene expression. DnaA regulates its own gene, and its steady-state level in the cell is determined by the cellular growth state. The frequency of replicon firing is dependent on the growth rate of the bacteria. As mentioned before, rapidly growing cells can have multiple copies of the genome, while cells with a very low growth rate have only one copy. Furthermore, as expected in cells with multiple genome copies, the genes near the origin will have a higher average copy number than the genes located near the terminus of replication and, therefore, will be more transcriptionally active.

In the case of multicopy plasmids, the control of copy number is mediated by the synthesis of *anti-sense* RNA of the replication initiator protein Rep, which is copied from the nontranscribed DNA strand and is thus complementary to the normal RNA. Anti-sense RNA prevents synthesis of the Rep protein, which is required for initiation of DNA synthesis and whose concentration is the primary mechanism of controlling initiation frequency. Rep proteins encoded by plasmids bind to additional copies of binding sites called "**iterons**," often present upstream of the *ori* sequences in the plasmids.

D. DNA Chain Elongation and Termination in Prokaryotes

Once initiated, DNA replication proceeds by coordinated copying of both leading and lagging strands. Although both bacteria and eukaryotes have multiple DNA polymerases, only one, named **polymerase III** (Pol III), is primarily responsible for replicative DNA synthesis in *E. coli*. In eukaryotes, DNA polymerases δ and ε have both been implicated in this process along with a suggestion that each of these two enzymes may be specific for leading or lagging strand synthesis.

Replication involves separation of two DNA strands which are catalyzed by DNA helicases which hydrolyze ATP during this reaction. ATP hydrolysis provides the energy needed for the unwinding process. All cells have multiple DNA helicases for a variety of DNA transactions. DnaB is the key helicase for replication of the genome *E. coli*. However, other helicases such as Rep and PriA are also involved in replication and interact with other components of the replication complex called the ***replisome***.

Replication requires a large number of proteins, including the holoenzyme of Pol III which includes, in addition to the catalytic polymerase cores, ten or more pairs of other subunits. The polymerase complex appears to have a dimeric asymmetric structure in order to replicate simultaneously two strands with opposite polarity. The continuous leading strand synthesis should be processive without interruption, because periodic RNA primer synthesis is not necessary once the leading DNA strand synthesis is initiated. On the other hand, the discontinuous lagging strand synthesis should not be processive, because repeated synthesis of RNA primers is required to initiate synthesis of each Okazaki fragment. The Pol III holoenzyme appears to assemble in a stepwise fashion, with its key β-subunit dimer acting as a sliding *clamp* based on its X-ray crystallographic structure of a ring surrounding the DNA. This clamp is loaded on DNA by the γ-complex, accompanied by ATP hydrolysis. The dimeric structure of the replication complex is maintained by the dimeric subunit of the holoenzyme. The β-clamp slides on the duplex DNA template and thus promotes processivity. Proliferating cell nuclear antigen (**PCNA**) is the sliding clamp homolog in eukaryotic cells and is also used in SV40 replication.

Much of the information about the composition of the *E. coli* Pol III holoenzyme, and DNA chain elongation, was generated from studies of the replication of small, single-stranded circular DNAs of bacterial viruses ϕX174 and M13 and also of **laboratory-constructed** plasmid DNA containing the *ori* (*ori C*) of *E. coli*. Asymmetric dimeric replication complexes have also been identified for larger *E. coli* viruses such as T4 with a linear genome and for the mammalian SV40 virus with a double-strand circular genome. In circular genomes, DNA synthesis is terminated at around 180° from the origin. In the case of linear genomes, termination occurs halfway between two neighboring replicons. The mechanism of termination is not completely understood. Although, in the *E. coli* genome, specific termination (***ter***) sequences are present, which bind to terminator proteins, such proteins act as anti-helicases to prevent strand separation. However, the termination may not be precise and occurs when the replicating forks collide.

E. General Features of Eukaryotic DNA Replication

Unlike the genomes in bacteria and plasmids (as well as in mitochondria and chloroplasts) which consist of a circular duplex DNA, with a single *ori* sequence, the genomes of

eukaryotes are not only much larger and linear, but also contain multiple *ori* sequences for DNA replication and thus multiple replicons. Thousands of replicons are simultaneously fired in mammalian genomes, as is needed to complete replication of the genome in a few hours. Mammalian genomes are three orders of magnitudes larger than the *E. coli* genome for which one round of replication requires about 40 min at 37°C. Replication of a mammalian genome, initiated at a single *ori*, would thus take more than 1 week with the same rate of synthesis. In fact, it would be even longer because the rate of DNA chain elongation is slower in eukaryotes than in *E. coli*, possibly because of the increased complexity of eukaryotic chromatin.

As mentioned earlier, DNA replication in eukaryotes occurs only during the S phase, which can last for several hours but whose duration varies with the organism, the cell type, and also the developmental stage. For example, in a rapidly growing early embryo of the fruitfly *D. melanogaster*, cellular multiplication with duplication of the complete genome occurs in less than 15 min. The details of temporal regulation of firing of different replicons are not known. However, euchromatin regions are replicated earlier than the heterochromatin regions.

The details of initiation of replication at individual replicons have not been elucidated in eukaryotes. Some *ori* sequences of the yeast genome, known as autonomous replication sequences (**ARS**), have been determined. Although such sequences in the mammalian genomes have not been isolated, the *ori* regions of certain genes which could be selectively amplified have been localized by two-dimensional electrophoretic separation. Nevertheless, a significant amount of information has been gathered regarding regulation of DNA replication at the global level.

F. Licensing of Eukaryotic Genome Replication

Unlike in bacteria and plasmids, DNA replication in eukaryotic cells is extremely precise, and replication initiation occurs only once in each cell cycle to ensure genomic stability. "**Licensing**" is the process of making the chromatin competent for DNA replication in which a collection of proteins called **origin recognition complex** (ORC) bind to the *ori* sequences. This binding is necessary for other proteins required for the onset of the S phase to bind to DNA. ORC is present throughout the cell cycle. However, other proteins required for replication initiation and chain elongation are loaded in a stepwise fashion. The onset of the S phase may be controlled by a **minichromosome maintenance** (MCM) complex of proteins which licenses DNA for replication, presumably by making it accessible to the DNA synthesis machinery. Several protein factors are involved in the loading process, which is regulated both positively and negatively. The level of regulator proteins, such as geminin, which blocks licensing, is also regulated by some cell cycle-dependent feedback mechanisms.

G. Fidelity of DNA Replication

The maintenance of genomic integrity in the form of the organism-specific nucleotide sequence of the genome is essential for preservation of the species during propagation. This requires an extremely high fidelity of DNA replication. Errors in RNA synthesis may be tolerated at a significantly higher level because RNAs have a limited half-life, even in nondividing cells, and are redundant. In contrast, any error in DNA sequence is perpetuated in the future, as there is only one or two copies of the genome per cell under most circumstances. Obviously, all organisms have a finite rate of mutation, which may be necessary for evolution. Genetic errors are one likely cause of such mutations. Inactivation of a vital protein function by mutation of its coding sequence will cause cell death. However, mutations that affect nonessential functions could be tolerated. Some of these mutations can still lead to change in the phenotype, which in extreme cases can cause pathological effects. In other cases, these may be responsible for susceptibility to diseases. In many cases, however, such mutations appear to be innocuous and are defined as an allelic polymorphism. The mammalian genome appears to have polymorphism in one out of several hundred base pairs. Such mutations obviously arose during the evolution and subsequent species propagation.

The error rate in replication of mammalian genome is about 10^{-6} to 10^{-7} per incorporated deoxynucleotide. The catalytic units of the replication machinery, namely, DNA polymerases, have a significantly higher error rate of the order of 10^{-4} to 10^{-5} per deoxynucleotide. In fact, some DNA polymerases, notably the reverse transcriptases of retroviruses, including HIV, the etiologic agent for AIDS, are highly error prone and incorporate a wrong nucleotide for every 10^2–10^3 nucleotides. These mistakes result in a high frequency of mutation in the viral protein, which helps the virus escape from immunosurveillance. The overall fidelity of DNA replication is significantly enhanced by several additional means. The editing or **proof-reading** function of the replication machinery is a $3' \rightarrow 5'$ exonuclease (which is either an intrinsic activity of the core DNA polymerase or is present in another subunit protein of the replication complex) which tests for base pair mismatch during DNA replication and removes the misincorporated base. Such an editing function is also present during RNA synthesis. In addition, after replication is completed, the nascent duplex is scanned for the presence of mispaired bases. Once such mispairs are marked by mismatch recognition proteins, a complex

mismatch repair process is initiated, which causes removal of a stretch of the newly synthesized strand spanning the mismatch, followed by resynthesis of the segment, as described later.

H. Replication of Telomeres—The End Game

Because DNA synthesis proceeds unidirectionally from $5' \rightarrow 3'$ with respect to deoxyribose, by sequential addition of deoxynucleotides to the 3′ terminus of the deoxynucleotide added last, chain elongation can proceed to the terminus of the template strand oriented in the 3′ to 5′ direction. But how about synthesis of the terminus of the complementary strand ? Because synthesis of this discontinuous (lagging) strand occurs in the opposite direction by repeated synthesis of a primer, the terminus could not be replicated. This problem of end replication is eliminated in the circular genomes of bacteria and the small genomes of plasmids and viruses. However, in the case of linear eukaryotic chromosome, the problem is solved by a specialized mechanism of telomere replication. **Telomeres** are repeats of short G-rich sequences found at both ends of the chromosomes (Fig. 6). In the human genome, the telomere repeat unit is 5′ (T/A)m Gn 3′, where $n > 1$ and $1 < m < 4$. **Telomerase** is a special DNA polymerase (reverse transcriptase) containing an oligoribonucleotide template 5′ Cn(A/T)m3′ (which is complementary to the telomere repeat sequence) as an integral part of the enzyme (Fig. 6). In the presence of other accessory proteins, telomerase utilizes its own template to generate the telomeric repeat unit and, by "slippage," utilizes the same oligoribonucleotide template repeatedly to generate thousands of repeats of the same hexanucleotide unit sequence. Because the lagging strand terminal region does not require an external DNA template, the newly synthesized DNA is present in an extended single-stranded region. Telomeres provide a critical protective function to the chromosome by their unique structures and prevent their abnormal fusion.

I. Telomere Shortening: Linkage Between Telomere Length and Limited Life Span

One profound implication of the specialized telomere structure and its synthesis is that in the absence of telomerase, the repeat length of telomeres could not be maintained. Telomerase is active in neonatal cells and also in some immortal tumor cells, but is barely detectable in diploid, terminally differentiated mammalian cells. Most such diploid cells can multiply *in vitro* in specialized culture medium, but have a limited life span. Loss of replicative capacity is associated with shortening of telomere repeat lengths. Furthermore, ectopic and stable expression of telomerase in human diploid cells by introduction of its gene confer an indefinite reproductive life on such cells. It is generally believed that cells will senesce if the telomere length is reduced below a critical level after repeated replication of the genome.

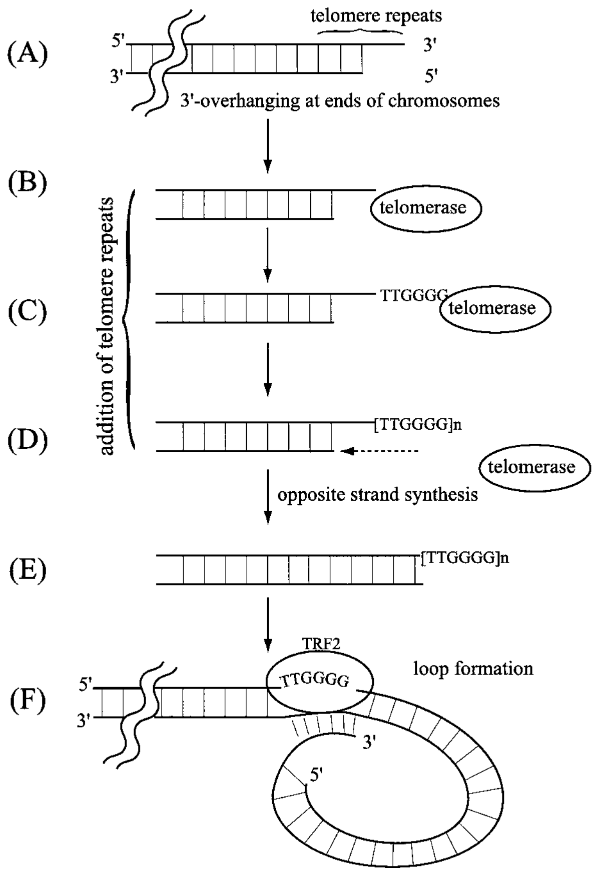


FIGURE 6 A schematic description of the role of telomerase in the maintenance of telomeres at chromosome termini. The double lines with break represent one telomere terminus of a chromosome in which the 5′ terminal region of the lagging strand is unreplicated (as in Fig. 4), resulting in an overhanging 3′ terminal region. In order to avoid shortening of this telomere sequence during successive rounds of replication, DNA template-independent telomerase extends the 3′ overhang by adding the telomere repeat sequence TTGGGG as shown in (**C**). The template for the repeat is an RNA present in the telomerase complex. The extended 3′ single-strand region then allows *de novo* initiation and filling in of the 5′ strand (**E**). Finally, the 3′ overhang loops to anneal with an internal sequence mediated by the telomere repeat factor (TRF2) in order to protect the terminus from degradation by nonspecific nucleases (**F**).

IV. MAINTENANCE OF GENOME INTEGRITY

The integrity of the genome, both in regard to sequence and to size, is essential for perpetuation of species. This integrity can be threatened in two ways. The first is by errors in DNA replication, as discussed earlier. A second inexorable process of DNA alteration occurs due to chemical reactions which can be either endogenous or induced by

external agents, including environmental genotoxic compounds, drugs, and radiation. Contrary to an earlier belief that DNA is a rather inert chemical, it is, in fact, sensitive to certain chemical reactions, e.g., depurination (loss of purine bases) and deamination of C to U, which occurs at a low but significant rate in DNA. It has been estimated that several hundred to several thousand such lesions are generated in the genome of a human cell per day. Both of these changes could be mutagenic. Loss of purines leads to **abasic sites** in DNA, which could direct misincorporation of wrong bases during DNA replication. Conversion of C into U is definitely mutagenic, because the change of a G•C to a G•U pair will give rise to one G•C pair and one A•T pair after DNA replication because U, like T, pairs with A. Often, C in the mammalian genome is methylated at the C-5 position, as discussed elsewhere, and 5-methyl C is deaminated more readily than C. Its conversion to T induces the same G•C $\rightarrow$ A•T mutation and, unlike deamination of C $\rightarrow$ U, does not produce an "abnormal" base in the DNA. A variety of environmental chemicals and both ultraviolet light present in sunlight and ionizing radiation from radioactive sources and X-rays induce a plethora of DNA lesions which include both base damage and sugar damage and are accompanied by DNA strand breaks. Many of these lesions, in particular, strand breaks and bulky base adducts, are toxic to the cells by preventing both replication and transcription. Other types of base damage and adducts can be mutagenic because they will allow DNA replication to proceed, but will direct incorporation of improper bases in the progeny strand.

A. Prevention of Toxic and Mutagenic Effects of DNA Damage by Repair Processes

Multiple repair processes have evolved to restore genomic integrity in all organisms ranging from bacteria to mammals. Excision repair comprises one class in which the damaged part of a DNA strand is excised enzymatically from the duplex DNA, leaving a single-strand gap. The gap is then filled by DNA polymerases starting at the 3′-OH terminus by utilizing the undamaged complementary strand as the template, followed by ligation of the nascent segment to the 5′ phosphate terminus at the other end of the gap with DNA ligase. The excision repair process consists of three subgroups which are utilized for distinct types of damage, although there is some overlap in their activities. **Base excision repair** is more commonly used for small base adducts, and **nucleotide excision repair** is used for replication/transcription-blocking bulky adducts. **Mismatch repair** evolved primarily to remove DNA mispairs that are generated as errors of replication. Both nucleotide excision and mismatch repair deficiencies have been linked to tumorigenesis, which results from mutation of critical oncogenes and/or tumor suppressor genes, thus causing uncontrolled cellular multiplication and prevention of cell death. Prevention of transcription of bulky adducts in active genes triggers nucleotide excision repair, at least in eukaryotes, in a process called "**transcription-coupled repair**." In fact, the repair complex has co-opted certain proteins of the transcription complex.

Although excision repair requires DNA synthesis, it is distinct from normal semi-conservative replication because it occurs throughout the cell cycle and may utilize nonreplicative DNA polymerases in both prokaryotes and eukaryotes. Pol II and Pol I in *E.coli* and DNA polymerase β have been identified as such repair polymerases. However, replicative polymerases can also be recruited in some cases, e.g., for mismatch repair synthesis.

Interestingly, during the last couple of years, a whole family of DNA polymerase have been identified and characterized in *E. coli*, yeast, and mammals (Table II). These enzymes are unique in their ability to bypass DNA base adducts which have lost the ability to base pair and thus are not utilized by standard DNA polymerases. It has been suggested that these replication bypass polymerases allow cell survival by allowing DNA replication even at the cost of introducing mutations.

B. Post-Replication Recombinational Repair

In contrast to the excision repair process in which the DNA damage is actually removed, both eukaryotic and prokaryotic cells have a novel mechanism of adapting to persistent, unrepaired damage by utilizing homologous **recombination** between the replicated progeny genomes. Recombination, the process of exchange between homologous DNA segments, involves unwinding of one duplex DNA and reciprocal strand exchange. When one strand in the parental DNA has a persistent lesion that prevents replication, a complete duplex is generated from the other, undamaged strand. The new strand subsequently acts as the template for the damaged region by strand exchange during replication of the damaged strand. Thus, recombination allows synthesis of the correct DNA sequence opposite the lesion.

V. DNA MANIPULATIONS AND THEIR APPLICATIONS

A. Episomal DNA and Recombinant DNA Technology

Extrachromosomal or episomal DNA, present in prokaryotes and lower eukaryotes, is distinct from the genome of organelles such as mitochondria or chloroplasts and serves many purposes. In bacteria, **plasmid DNA** can be

transmitted to progeny cells, and the genes in these plasmids encode distinct proteins which provide growth advantage or survival to the host bacteria. For example, many proteins which confer drug resistance by a variety of mechanisms are encoded by the plasmids, which are invariably present as double-stranded circular DNA containing several to hundreds of kilobase pairs.

The plasmid DNAs are self-replicating genomic units which are completely dependent on the host bacteria or yeast for their replication. These are also critical vehicles for **recombinant DNA technology** based on cutting and rejoining DNA fragments. Its invention, some three decades ago, revolutionized molecular biology and is at the root of nearly all modern breakthroughs in biology. **Restriction endonucleases**, which are enzymes characterized by stringent recognition of specific DNA sequences, cleave DNA duplexes and often leave identical terminal sequences in both plasmid DNA and a gene or segment of a genome. The fragments can then be joined by a DNA ligase. Joining heterologous fragments generates recombinant DNA, for example, a circular plasmid molecule containing foreign genes. These DNA molecules can then be introduced into living cells which allow their reproduction, so that a large amount of recombinant plasmid can then be generated.

Recombinant plasmids specific for bacteria, yeast, and even mammalian cells have been generated in the laboratory and exploited for a variety of basic and applied research applications. Specifically, recombinant **expression plasmids** can be constructed in order to express the ectopic protein encoded by the foreign (*trans*) gene in the appropriate host cell. Recombinant plasmids of mammalian cells are based on viruses, rather than on episomal DNA. Only the DNA replication function of the virus is incorporated into the plasmid, so that the plasmid is replicated without producing the active virus. In the case of human cells, simian virus 40 (SV40) is commonly used to generate recombinant DNA.

The circularity of the plasmid is essential for *E. coli*, but not mammalian or yeast cells. This may be consistent with the circular genome of the bacteria vs linear genomes of eukaryotes. However, plasmid vectors specific for mammalian cells must be propagated, preferably in *E. coli*. Such "shuttle" vectors are therefore required to have a circular configuration.

B. Polymerase Chain Reaction (PCR)

A critical advance in molecular biology came with the invention of PCR, based on a remarkably simple principle, and revolutionized many important aspects of biomedical research and medical jurisprudence. The method is based on the rationale that each strand of a piece of DNA sequence can be replicated repeatedly by using an oligonucleotide primer and a DNA polymerase (Fig. 7). After a duplex DNA molecule is generated, the next cycle is carried out by separating the two strands by heating and then starting the next cycle of synthesis after annealing oligonucleotide primers to each template strand. Thus, the repeated cycles of synthesis, denaturation, and primer annealing to both strands allow synthesis of a specific DNA sequence at an exponential rate. Thus, a tiny piece of a DNA molecule could be amplified about a millionfold after 20 cycles of this chain reaction (assuming 100% efficiency of the process; Fig. 7).

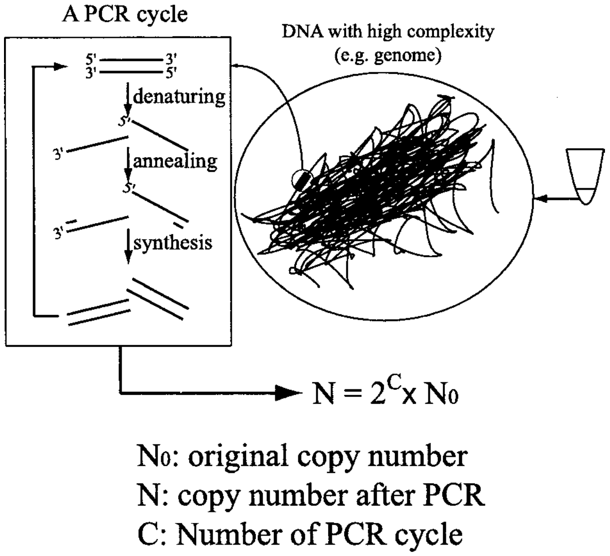


FIGURE 7 Principle of polymerase chain reaction (PCR). A copy of a relatively short fragment of DNA (0.1–20 kilobase pairs) can be specifically amplified from genomic DNA by PCR. A typical PCR reaction mixture contains genomic DNA; two oligonucleotide (~20 bp) primers, which have same sequences as the two ends of the DNA fragment to be amplified; and a thermostable DNA polymerase. A cycle of PCR reaction consists of three steps, starting with denaturing the genomic DNA at high temperature (e.g., 95°C), followed by primer annealing at near Tm (melting temperature for primer-DNA hybridization), followed by DNA synthesis from the primers by the DNA polymerase. Theoretically, the copy number of the DNA of interest (N) can be amplified to $2^C \times N_O$, where N_O is the original copy number and C is the number of PCR cycles.

The PCR technology became viable after discovery of thermostable DNA polymerases derived from bacteria, such as *Thermobacillus aqualyticus* (**Taq**), which grow at high temperature. The cycles of PCR could then be automatically set in a thermal cycler. PCR does have some limitations. The most important of these are: (1) errors in DNA replication; (2) less than complete efficiency in each step of the reaction; and (3) improper primer annealing when complex DNA is used. Thus, when amplification of a segment of DNA in a complex genome is desired, the

first requirement is the sequence information for the termini of the segment, based on which the oligonucleotides will be designed for each terminus and then synthesized. However, errors of replication cannot be completely eliminated. Any error in DNA synthesis that occurs early will be perpetuated. Furthermore, if replication is initiated by primers annealed to an incorrect DNA sequence, the wrong PCR product will be generated.

Primarily, because it has both sensitivity and specificity, PCR technology has revolutionized many aspects of biomedical research. Several modifications of the basic methodology have provided additional powerful tools. For example, a trace amount of RNA can be quantitated by reverse transcriptase PCR (RTPCR), where a reverse transcriptase synthesizes the complementary DNA strand of the RNA, which then serves as the template for regular PCR.

DNA in a very small amount of biological samples can be amplified by PCR. This technique has been exploited in criminal investigations to identify suspects by "**fingerprinting**" their DNA, which involves determining a characteristic pattern of repeat sequences in the genome after PCR amplification of the total DNA. PCR has also been utilized in the identification of pathogens and other microorganisms, based on certain unique sequences of each organism. PCR has been exploited for a variety of *in vitro* manipulations of DNA sequences in plasmids, viruses, and synthetic DNA by generating site-specific mutations and a variety of recombinant DNA plasmids.

VI. TRANSCRIPTIONAL PROCESSES

Transcription is a highly complex process because of its defined initiation and termination sites in the genome and the subsequent processing and regulation of its synthesis. The steady-state level of a protein in the cell is the balance of its rate of synthesis and degradation. The synthesis is determined primarily by the steady-state level of its mRNA. Thus, the rate of transcription often determines the level of its gene product *in vivo*.

As mentioned earlier, RNA synthesis is catalyzed by the RNA polymerase in all organisms. Prokaryotes express a single RNA polymerase used for synthesis of all RNAs, while eukaryotes encode multiple RNA polymerases with dedicated functions. RNA polymerase I (Pol I) in eukaryotic cells is responsible for synthesis of ribosomal RNA, which accounts for more than 70% of total RNA in the cell. Pol III catalyzes synthesis of small RNA molecules, including transfer RNAs which bring in appropriate amino acids to the ribosome for protein synthesis by using their "anti-codon" triplet bases. Pol II is responsible for synthesis of all other RNA, specifically mRNA.

RNA polymerases of all organisms are complex machines consisting of multiple subunits which alter conformation. A variety of structural analyses show the presence of a 2.5-nm-wide "channel" on the surface of all DNA polymerases which could be the path for DNA. The RNA polymerase holoenzyme binds to a promoter-specific recognition sequence upstream ($5'$ side of the transcribed strand) of the site of synthesis initiation. While the RNA polymerase is normally present as a closed complex with nonspecific DNA, in which DNA base pairs are not broken, a significant conformational change produces the **open complex** when RNA the enzyme binds the promoter, unwinds the DNA duplex, and is poised to initiate RNA synthesis.

As in the replication process, ***initiation*** is the first stage in transcription and denotes the formation of first phosphodiester bond. Unlike in the case of DNA synthesis, RNA chains are initiated *de novo* without the need of a primer. However, when a primer oligonucleotide is present, RNA polymerases can also extend the primer as dictated by the template strand. A purine nucleotide invariably starts the RNA chains in both prokaryotes and eukaryotes, and the overall rate of chain growth is about 40 nucleotides per second at 37°C in *E. coli*. This rate is much slower than that for DNA chain elongation ($\sim$800 base pairs per second at 37° for the *E. coli* genome).

RNA synthesis is not monotonic, and RNA polymerases can move backward like DNA polymerases do for their editing function in which an incorrectly inserted deoxynucleotide is removed by $3'$ exonuclease activity. RNA polymerases stall, back track, and then cleave off multiple newly inserted nucleotides at the $3'$ terminus. Subsequently, polymerases move forward along the DNA template and resynthesize the cleaved region. Based on the segment of DNA covered by an RNA polymerase as analyzed by **DNA footprinting**, it has been proposed that the enzyme alternatively compresses and extends in its binding to the DNA template and acts like an inchworm in its transit.

RNA polymerases of both prokaryotes and eukaryotes function as complexes consisting of a number of subunits. The *E. coli* RNA polymerase enzyme with a total molecular mass of about 465 kD contains two α-subunits, one β- and one β'-subunit each, and a σ-subunit which provides promoter specificity. During chain elongation, a ternary complex of macromolecules among DNA template, RNA polymerase, and nascent RNA is maintained in which most of the nascent RNA molecule is present in a single-stranded unpaired form. The stability of the complex is maintained by about nine base pairs between RNA and the transcribed (noncoding) DNA strand at the growing point.

While DNA replication warrants permanent unwinding of the parental duplex DNA, asymmetric copying of only

one strand by RNA polymerase requires localized strand separation which is induced by the polymerase itself, resulting in a **transcription bubble**. During chain elongation, this bubble moves along the DNA duplex. Initiation of RNA synthesis is enhanced in an *in vitro* reaction with supercoiled duplex circular DNA template in which base pairs are destabilized due to torsional stress. Unwinding of the helix at the transcription site causes overwinding (positive supercoiling) of the template DNA ahead of the transcription bubble and underwinding (negative supercoiling) behind the bubble.

A. Recognition of Prokaryotic Promoters and Role of σ-Factors

In prokaryotic RNA polymerases, the σ-factor is required for promoter recognition and binding. It is loosely bound to the core complex and released after the nascent RNA chain becomes 8–9 nucleotides long. The core polymerase with σ-factor has a high affinity for nonspecific DNA. The σ-factor alters the conformation of the holoenzyme so that its affinity for nonspecific DNA is reduced and the specific binding affinity for the promoter is significantly enhanced.

More than one type of σ-factor is present in *E. coli*, and more such factors are present in other bacteria. These different factors may have specialized functions in altered growth conditions, cause a global change in transcriptional initiation due to their recognition of distinct −35 and −10 sequence elements, and have a preference for different promoters.

RNA chain termination in bacteria occurs by two mechanisms, one with assistance of a protein factor rho (ρ) and the other without need of a protein. In both cases, termination occurs at a specific **terminator** sequence in the gene, at which the RNA polymerase stops adding nucleotides to the growing RNA chain, which is then released from the template. The terminator sequence often has a "**hairpin**" structure which results from intramolecular base pairing in a palindromic sequence. It is likely that such hairpins at the end of RNA promote its dissociation from DNA. Termination can be prevented by an anti-terminator protein, which allows the polymerase to ignore the terminator signal.

A unique distinction between prokaryotic and eukaryotic RNA synthesis is the temporal relationship between its synthesis and utilization in information transfer. Prokaryotic transcription of mRNA is linked to its reading on the ribosome for protein synthesis. Thus, even before transcription is terminated, the 5′ terminal region of the nascent mRNA is complexed with a ribosome for initiation and propagation of protein synthesis. In the case of eukaryotes, transcription occurs in the nucleus, from which the RNA has to be transported to the endoplasmic reticulum with ribosomes in the cytoplasm. Two sequence motifs that are common constituents of promoters in prokaryotic genomes and are nominally referred to as −35 and −10 sequences signify that the midpoint of these sequences are located 35 and 10 bp 5′ of the start site of transcription. However, the exact distance is somewhat variable for different genes. The consensus −35 sequence is TTGACA, and the consensus of −10 is TATAAT. However, both of the sequences are also somewhat variable. The strength of a promoter, i.e., how efficiently it is recognized for transcriptional initiation, depends on the exact sequence of the −35 and −10 sequences and possibly the intervening "spacer" sequences as well. The promoter strength can vary widely among genes, and mutations in the −35 or −10 sequence in a particular gene can dramatically affect its promoter strength.

B. Regulation of Transcription in Bacteria

Unlike replication of the complete genome, which is essential for cellular propagation, not all genes need to be transcribed in a particular cell for its survival. Synthesis of mRNA is required for generation of proteins. Because not all proteins are required at all times for cellular survival and metabolism, both in prokaryotes and eukaryotes, and many proteins are expressed only in specific stages of development and differentiation in higher eukaryotes, a gene's transcription is often highly regulated. Furthermore, the stability of mRNAs and the proteins they encode vary over a wide range. Thus, different mRNAs are not made at the same rate. Additionally, the bulk of RNA, and in fact a large fraction of the cell mass, consists of ribosomal and transfer RNAs needed for carrying out protein synthesis. Both ribosomal and transfer RNAs are extremely stable.

Regulation of transcription, first investigated in bacterial viruses, primarily in *E. coli*, an intestinal microbe and its bacteriophage λ, is the foundation of molecular genetics. The ease of generating and manipulating mutants of various genes in *E. coli* and λ led to the discovery of **repressors**, which are proteins that bind to operator sequences of genes and turn off transcription. The genes that were originally studied encode enzymes for sugar (lactose and galactose) metabolism. Inactivation of these genes and their expression could be studied because the proteins are not essential for bacterial survival. An activator needed for expression of lactose-metabolizing β-galactosidase was identified; it is downregulated in the presence of glucose ("glucose effect") and upregulated by binding to 3′-5′ cyclic AMP.

Significant advances in elucidating the mechanism of transcriptional regulation came from the life cycle studies of the lysogenic λ virus, whose virus-specific proteins are not expressed in the lysogenic state, when its duplex DNA

genome is linearly integrated in the host chromosome. Here again, both positive and negative regulatory mechanisms are in play to fine tune the expression of genes from a low maintenance level during lysogeny to large-scale expression of the viral genome when the lysogenic virus enters the lytic phase of growth and exploits the host cell synthetic machinery for replication of its own viral DNA, RNA, and proteins.

C. Eukaryotic Transcription

The fundamental process is identical in prokaryotes and eukaryotes, in that an RNA polymerase complex binds to the promoter and initiates transcription at a start site downstream to the promoter. *De novo* initiation of an RNA chain occurs with a purine nucleotide and creation of a transcription bubble with the open complex. The transcription complex can slide back along the nascent chain and endonucleolytically cleave off the 3′ segment, then moves forward along the DNA template chain; termination occurs at specific regions in the genes.

In spite of this similarity, however, the details are very different in eukaryotic cells and are summarized as follows.

1. Eukaryotic RNA polymerases contain many more subunits, located in the different regions of the nucleus. Pol I, specific for synthesizing rRNA, is located in the nucleolus, a specialized structure within the nucleus, while Pol II and Pol III are in the nucleoplasm. These enzymes have 8–14 subunits with a total molecular mass >500 kD. The large subunits have some sequence similarity with the bacterial RNA polymerases. RNA polymerases of mitochondria and chloroplasts are phylogenetically closer to bacterial RNA polymerase, commensurate with the fact that the target genes of these enzymes are fewer and much smaller in organelles, which are thought to have arisen by symbiotic acquisition of bacteria by primitive eucaryotes.

2. The promoter composition and organization of eukaryotic polymerases are quite specific for each polymerase. The promoters of rRNA genes contain a core and an upstream control element which is needed for high promoter activity. Two ancillary factors, UBFl and SLl, bind to these sequences. Although SLl binds only after UBFl in a cooperative fashion, SL1 is a σ-factor with four proteins among which TBP is also required for initiation by the other polymerases. Pol I is akin to Pol III in that it utilizes both upstream and downstream promoters. There are two types of internal promoters with distinct sequence boxes. One **transcription factor** (TFIII B) is required for initiation of RNA synthesis by Pol III. Other factors (TFIII A and TFIII C) help TFIII B bind to the right location and act as **positioning** factors for correct localization of Pol III initiation.

Pol II is the most versatile and widely utilized RNA polymerase *in vivo* and absolutely needs auxiliary, transcription factors (TFII) whose requirement is dependent on the nature of promoters.

3. The nature of eukaryotic promoters is quite different from the prokaryotic promoters. In addition to the bipartite promoter of Pol I, both Pol II and Pol III have a "TATA box" located about 25 bp upstream of the start site in Pol II responsive genes. The 8-bp sequence consists of only A•T base pairs and is surrounded by G•C pair-rich sequences. Interestingly, the TATA box is quite similar to the −10 sequence in *E. coli* promoters.

There is a second element called a CAAT box, usually about −15 bp 5′ of the TATA box. Alternatively a G•C rich sequence is present in some promoters, often at position −90. The consensus GC box sequence is GGGCGG, of which multiple copies are often present and occur in both orientations. These elements are not all present in all promoters; it appears that they work in a "mix and match" fashion. These boxes, and also a octamer box, bind to specific *trans*-acting factors and are engaged in multiple protein interactions among themselves as well as with components of the RNA Pol II holoenzyme.

There is no significant homology among transcription start sites of various genes, except for the tendency for the first base in the transcript to be an A flanked on either side by pyrimidines. This region is defined as the **initiator**.

The first step in transcriptional initiation of a TATA-containing promoter is the binding of the factor TFIID to the region upstream of the TATA site. The TATA-binding protein, TBP, which specifically binds to the TATA box, is a component of the TFIID complex, along with other proteins collectively called TAFs (TBP-associated factors). TAFs can be variable in the TFIID complex, both in species and amounts, and provide the promoter specificity for initiation. Some TAFs are tissue specific. TFIID has a molecular mass of 800 kD, containing 1 TBP and 11 TAFs. TBP acts as a positioning factor and is able to interact with a wide variety of proteins, including Pol II and Pol III. It binds to the minor groove of the DNA double helix and makes contact with other factors which mostly bind to the major groove and can make multiple contacts. By bending the DNA at the binding site, it appears to bring the factors and RNA polymerase into closer proximity.

Although TBP is utilized by both Pol II and Pol III, TFIID is the specific complex for Pol II recognition of a promoter. Other transcription factors (e.g., TFIIA) bind to the TFIID promoter complex and cover increasing segments of DNA. In addition to TFIIA, these include TFIIE, TFIIF, TFIIH, and TFIIJ. Most of the TFII factors

are released from the transcription complex before Pol II leaves the promoter and carries out chain elongation. Interestingly, the same general transcription factors, including TFIID, bind to the TATA-less promoter, even though TATA binding by TBP is not available.

There is an important contrast in the assembly of RNA polymerase complexes in eukaryotes and prokaryotes. *E. coli* RNA polymerase binds to the promoter as a complex with the σ-factor, providing the specificity for initiation but not elongation. Eukaryotic Pol II, on the other hand, goes through a much more complex choreography because of the prerequisite for binding to the promoter by other transcription factors. This dichotomy reflects the complex structural organization of the eukaryotic genome and the presence of a much larger number of genes with their complex regulation. Such regulation is not only dependent on the environment, but also on the stage of development and differentiation, at least in the metazoans.

4. A unique difference between prokaryotic and eukaryotic transcription is that in prokaryotes a single mRNA containing many genes can be transcribed from the DNA template as a single transcription unit, coupled with their direct translation on ribosomes into discrete polypeptides. This process reflects the fact that genes which encode enzymes in a given pathway are often clustered in an operon and are co-ordinately regulated.

In contrast to the synthesis of polycistronic mRNA in *E. coli* and other bacteria, eukaryotic transcription units usually consists of single genes. This characteristic may also reflect uncoupled transcription and translation in these organisms. Thus, heterogeneous nuclear RNA (hnRNA) is synthesized in the nucleus and then transported to the cytoplasm along with its processing into mature mRNA including splicing, addition of poly(A) tail at the 3′ end, and capping at the 5′ end. Subsequently, the RNA is translated on ribosomes (endoplasmic reticulum). Thus, synthesis and utilization of mRNA are temporally and spatially separated.

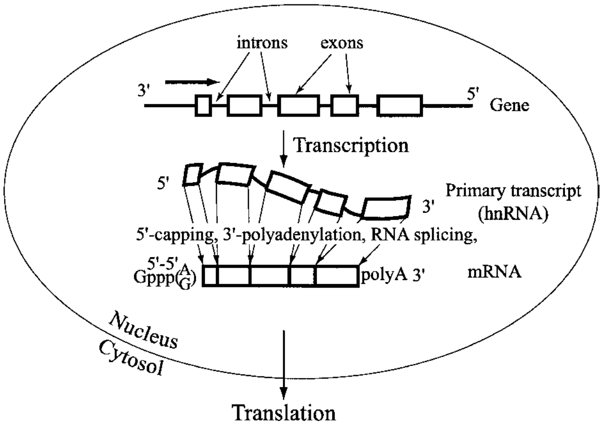


FIGURE 8 A schematic representation of RNA splicing. The coding sequence in metazoan genomes is usually present in segments (exons; indicated by boxes) interspersed between noncoding introns. After synthesis of the primary RNA transcript (called heterogeneous nuclear RNA or hnRNA), the intron sequences are removed by precise cleavage and rejoining is mediated by the spliceosome complex, so that the resulting mature mRNA contains a correctly juxtaposed coding sequence for the polypeptide. The mRNA is also "capped" by 5′-5′ linkage with GMP, and a tail of poly(A) is added at the 3′ terminus to increase the stability of mRNA and to enhance its efficiency in directing protein synthesis when the mRNA is transported from the nucleus to the cytoplasm.

D. RNA Splicing in Metazoans

The central dogma of molecular biology that the information flow from DNA to RNA to protein involves colinearity of the sequences of the monomer units is somewhat violated in metazoans because of the presence of interrupted or fragmented genes (Fig. 8). Thus, while the polypeptide sequence is colinear with the codons of the coding sequence in the mRNA, the RNA itself is not collinear with the gene from which it is transcribed. In other words, the gene contains additional intervening sequences called introns, which are transcribed but whose RNA sequence is subsequently removed from the final mRNA containing the coding sequence. The primary gene transcripts of nuclear genomes, called **heterogeneous nuclear RNA** (hnRNAs), are present in a form of protein-bound particles (**ribonucleoprotein particles**, or hnRNP). RNA splicing is the process of excising introns from hnRNAs, and contiguous exons are then joined to form mature mRNAs, which are subsequently translocated to cytoplasm and are used as templates for translation (Fig. 8). The cleavage and rejoining occur at specific junctions between exons and introns, so that there are no errors in mature mRNA. First, two adjacent exons are aligned, while the intervening intron is extruded, forming a loop ("lariat") structure. Then the upstream exon is cleaved and joined to the downstream exon via a transesterification reaction. In most cases, two factors are essential for this process. One, the *cis*-elements in introns and exons, is the signaling sequences for the exact junction sites. The other is the splicing machinery, consisting of several small ribonucleoprotein particles (snRNP; U1, U2, and U4–U6), each of which contains small RNA molecules and proteins. The U1 and U2 snRNPs contain RNA complementary to the intron *cis*-element and catalyze the formation of the intron lariat, while two adjacent exons are aligned together. With other snRNPs forming an intermediate complex (**spliceosome**), U6 catalyzes the transesterfication. It should be noted that introns in RNA of some lower eukaryotic species are autospliced and therefore do not require snRNPs.

Termination of eukaryotic transcription is coupled with processing. The mature rRNA is obtained by cleavage of a larger primary transcript synthesized by Pol I. Termination of Pol II transcription occurs at a repeat sequence of U, as in the case of *E. coli* RNA polymerase, but without the presence of a hairpin structure. More importantly, the 3′ termini of mRNAs are generated by cleavage of primary precursor transcripts followed by addition of a **tail** of poly(A), a homopolymer of up to several hundred AMP residues synthesized by poly(A) polymerase in a template-independent reaction.

E. Regulation of Transcription in Eukaryotes

While both prokaryotic and eukaryotic genes are regulated by activators and repressors, enhancer elements are unique to eukaryotic genes and can profoundly increase the rate of transcription. These elements are located at a variable distance from the basic promoter itself, can be present both upstream or downstream to the promoter, and, in fact, can even be within the transcription unit. One unexpected feature is that they can function in either orientation and can activate any promoter located in the vicinity.

Upstream activating sequences (UAS) have been identified in yeast and are analogous to enhancers in the mammalian genes. Based on the known properties of enhancers, it appears that the presence of these sequences affects chromatin structure and/or the helical structure of the DNA template itself. Further studies are needed to test other possibilities as well, e.g., whether the enhancer provides an entry point for the transcription complex or is needed to place the template at the nuclear matrix where transcription takes place.

Positive and negative regulation of prokaryotic genes is achieved by binding of activators and repressors, respectively, to their cognate binding sites in the genes. Down-regulation is more common, at least in *E. coli*, than positive regulation. In fact, the same protein can provide dual functions in a few cases, depending on the location of the sequence motif.

In contrast, because of the complexity of chromatin structure and genomic organization development, differentiation, and cell cycle-specific synthesis of proteins, regulation of eukaryotic genes is extremely complex. This is evident from the large number of families of regulatory *trans*-acting factors which recognize similar if not identical sequence motifs in different genes. Sometimes, these factors have a distinct modular structure—one module for binding to target DNA sequence and another for interaction with components of the transcription apparatus.

On top of these complexities, the signal for initiation of transcription may be extracellular, e.g., a growth factor which induces cell proliferation. A highly complex signaling cascade is initiated in response to the first signal. The external ligand first binds to its receptor on the cell surface, followed by internalization of the receptor ligand complex. A series of reversible chemical modification (mostly phosphorylation of the regulatory proteins) finally activates the ultimate transcription factors, which then trigger transcription of target genes.

The unique difference between the eukaryotes and prokaryotes is in the utilization of transcription factors. In bacteria, one factor is usually specific for one gene or one regulatory unit. In eukaryotes, on the other hand, a single factor activates multiple target genes.

Prokaryotic regulatory processes have been elucidated in remarkable detail by utilizing the power of molecular genetics, including "reverse genetics" by which the chromosomal genes in the organism could be mutated at specific sites and the mutant gene products purified and characterized. Furthermore, these genes can be expressed in the episomal state by introducing them into autonomously replicating recombinant plasmids.

Commensurate with the significantly higher complexity and size of the genome and differentiation and developmental stages in metazoans, gene regulation in these organisms is very complex and occurs at many levels. Sets of genes are activated at distinct stages of differentiation and development of multicellular organisms in order to encode proteins which are required for specialized functions of the cells in these stages. In contrast, certain "housekeeping" proteins, including enzymes for metabolism and synthesis of all cellular components (i.e., RNA, DNA, structural proteins, and lipids), as well as enzymes for biosynthetic and degradative pathways, are needed in all cell types and developmental stages. Most somatic cells in adult mammals are nondividing and therefore do not require DNA synthesis machinery. However, all cells require transcription for generating proteins for other cellular functions. Unraveling the molecular mechanisms of regulation is the major focus of current research in molecular biology. The regulatory process is affected by multiple parameters.

Many genes are activated due to external stimuli, e.g., exposure to hormones and growth factors. In these cases the extracellular signal often acts as a ligand to bind to cell surface receptors which activate the *trans*-acting factor(s) via multiple steps of signal transduction.

1. Regulation of Transcription via Chromatin Structure Modulation in Eukaryotes

The eukaryotic genome is organized at multiple levels, starting with the nucleosome core as described earlier. The nucleosomes are organized in a higher order chromatin structure due to increasing compaction of DNA: from

2-nm-wide naked DNA fiber to metaphase chromosomes of microscopic width. The DNA template has to be accessible to transcription machinery containing RNA polymerase; transcriptionally inactive, highly compacted chromatin maintains its structure by multiple protein–protein and protein–DNA interactions, which are yet to be elucidated. However, it is now clear that at the nucleosome level, it is the strength of interaction between histones and DNA which regulates accessibility of the DNA to the transcription machinery, a process controlled by acetylation and phosphorylation of core histones. Multiple histone acetylases and deacetylases, which are themselves regulated, modulate chromatin structure. As stated previously, large protein complexes named SWI and SNF modulate chromatin structure in an energy-dependent process which may be responsible for the differentiation/development-dependent turning on or off of specific sets of genes.

2. CpG Methylation-Dependent Negative Regulation of Genes

In addition to histone modification, DNA itself was found to be modified, most commonly by methylation at the C-5 position of cytosine, but only when it is present as a CpG dinucleotide. Such methylation, catalyzed by specific methyltransferases, invariably inhibits gene expression, which was unequivocally established in the genomes during embryonic development. Sets of genes are selectively methylated or demethylated in the CpG sequences, most commonly in the genes' promoter regions, leading to their activation or repression. Proteins that bind to methylated CpG sequences have been implicated in the control of histone deacetylation, thereby leading to closing of the promoter.

F. Fidelity of Transcription (RNA Editing)

The informational content of gene transcripts can be altered during or after transcription by a process collectively called RNA editing. The information changes are carried out at the level of mRNA. RNA editing appears to be a widespread phenomenon for both normal and aberrant RNA processing in organelles and nuclei. It was first discovered in the mitochondria of kinetoplasts in protozoa. Two types of RNA editing have been observed: (1) alteration of coding sequence by nucleotide insertion and/or deletion and (2) base substitution. In mammalian cells, editing of an individual base in mRNA can cause a change in the sequence of the protein. Such changes can occur by enzymatic deamination in which C is converted to U or A is converted to hypoxanthine. Change of U to C has also been observed in many plants. The (mitochondrial) mRNAs of several kinetoplastid species (*Crithidia*, *Trypanosoma*, etc.) were found to be edited by the insertion and deletion of U's at many sites in mRNAs. The editing process uses a template consisting of a guide RNA (gRNA) whose genes function as independent transcription units. The gRNAs are generally 55–70 nucleotides in length and complementary to the mRNA for a significant distance including and surrounding the edited region. The gRNA dictates the specificity of uridine insertions by its pairing with the pre-edited RNA, but also provides the U residues that are inserted into the target RNA by transesterification reactions; the reaction proceeds along the pre-edited RNA in the 3′-5′ direction. The RNA editing process reveals the existence of a previously unrecognized level for the control of gene expression. Recognition of this process has resulted in an expansion of the central dogma. Multiple RNA editing processes play a significant role in normal physiological processes, as well as being responsible for some disease.

VII. CHEMICAL SYNTHESIS OF NUCLEIC ACIDS (OLIGONUCLEOTIDES)

Development of strategies for chemical synthesis of nucleic acids represented a major breakthrough in molecular biology, because most of the current approaches involving PCR, manipulation of recombinant DNA, studies of gene regulation, etc. require synthetic DNA and RNA oligonucleotides with defined sequences. The difficulty of synthesizing RNA and DNA polynucleotide chains from mononucleotide units lies in the reactivity of the side chains of the bases and the susceptibility of the sugar glycosyl bond to cleavage under the harsh conditions needed for condensation reactions to generate phosphodiester bonds. An additional problem in RNA synthesis is the presence of the C′2-OH group in ribose.

H. Khorana's group was the first to solve the problem by blocking all reactive side chains of the bases with reversible blocking groups; a phosphodiester bond between C′3-OH of one nucleotide and the C′5-phosphate of another was generated by condensation in the presence of dicyclohexyl carbodiimide (DCC) under mild conditions. Repeating the process in a cyclic fashion generated oligonucleotides of a defined sequence. While the DCC condensation was efficient, the whole process was extremely laborious, because the products of each reaction had to be purified free of the side products and the blocking groups had to be removed after each cycle. Furthermore, the efficiency of the synthetic reaction fell off rapidly with increasing size of the oligonucleotide.

A major advance occurred in the 1970s when two distinct types of chemistries were invented for synthesis of

deoxyoligonucleotides with the possibility of automating the cyclic procedure. One was based on phosphodiesters of deoxynucleotides as the starting material, which had been utilized early on for synthesis of oligodeoxynucleotides. However, the phosphoramidite method invented later has become the exclusive method of choice for synthesis of both RNA and DNA sequences. The advantages of this method are (1) the relatively high stability of the starting compounds and (2) the mild reaction conditions for removal of the protective groups.

Automated procedures have been developed for solid-state synthesis of polymers (Fig. 9), which is initiated by covalent attachment of the first monomer phosphoramidite unit to a glass matrix in the reaction vial; the phosphodiester condensation reaction is carried out by addition of monomer units in the $3' \rightarrow 5'$ direction, which is opposite to the direction of enzymatic synthesis. Each cycle of synthesis involves removal of the protective groups after the condensation reaction. Fixed amounts of phosphoramidites of four nucleotides, as well as other modified nucleotides, are added to the reaction vial in predetermined order and amounts. The chemical treatments involving acidic and alkaline solvents are carried out in a preprogrammed sequential order, and the glass matrix containing the oligonucleotide is washed with solvent in between each reaction. The complete procedure has been automated in several commercial instruments. After synthesis is completed, the oligonucleotide product is released from the glass matrix by alkaline treatment, and then the last protective trityl group is removed. The quality and efficiency of polymer synthesis is determined by the efficiency of the individual reactions. The major advantage of phosphoramidite-based synthesis is very high efficiency (99%) of both the condensation and the deprotection reactions. Nonetheless, it is obvious that because the final yield of the oligonucleotide is the product of the yields of each individual cycle, very long oligonucleotides cannot be synthesized at a significant level. In practical terms, the current size limit of an oligonucleotide is usually up to about 120 monomer units. Even then the product has to be purified (usually by gel electrophoresis) from the contaminants, mostly composed of failed synthesis material.

A major problem in therapeutic use of oligonucleotides is their degradation by nonspecific nucleases, once delivered inside the tissues and cells. One of several approaches to counter this problem is to synthesize artificial nucleic acids in which phosphate oxygen is replaced with sulfur. In a phosphorothioate oligo (S-oligo), some or all of the internucleotide phosphate groups are replaced by a phosphothioate group. These S-oligos are widely used

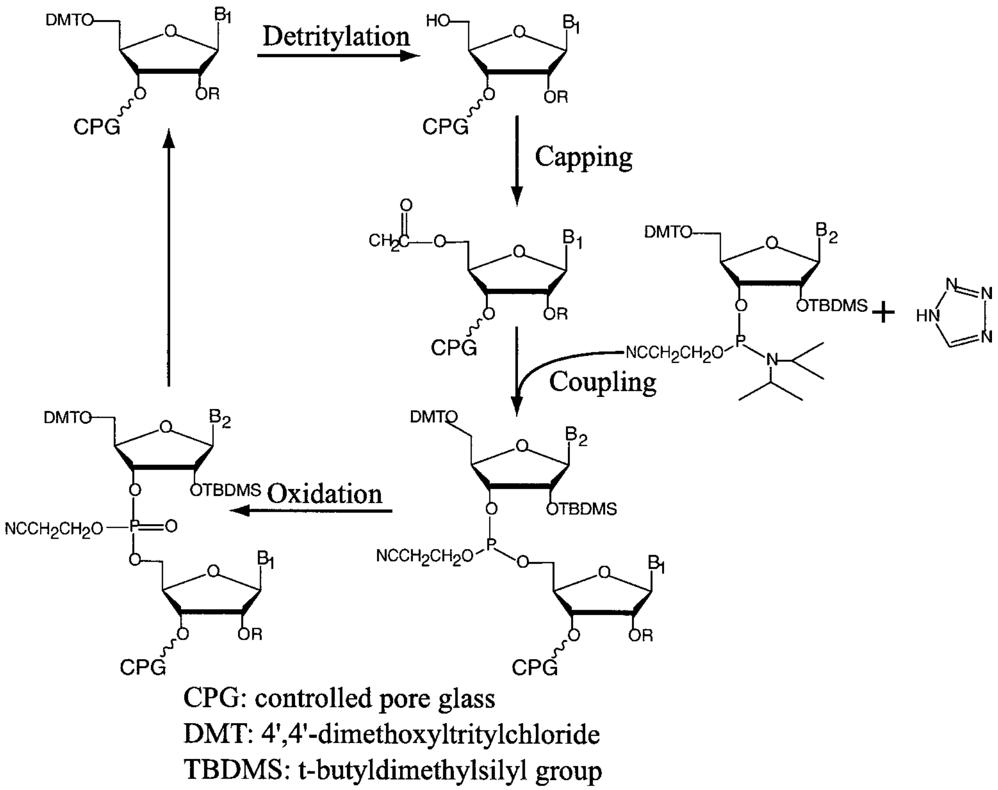


FIGURE 9 An outline of the chemical synthesis of nucleic acids.

in anti-sense applications because of their enhanced stability. The modified backbone of an S-oligo is resistant to the action of most nucleases and endonucleases, but they also tend to be subject to more nonspecific interactions due to “stickiness.”

A. Peptide Nucleic Acids (PNA)

Peptide nucleic acids (PNA; Fig. 10) are synthetic polynucleobase molecules which bind to DNA and RNA with high affinity and specificity. PNA was constructed with a charge-neutral, achiral, pseudopeptide backbone and is therefore chemically more closely related to peptides than to nucleic acids. Thus, PNAs, because of their backbone properties, show extremely good nucleic acid hybridization properties. In fact, PNA–DNA and PNA–RNA duplexes are, in general, thermally more stable than the corresponding DNA(RNA)–DNA(RNA) duplexes.

PNAs are relatively easy to synthesize and are stable (especially biologically). These make PNA an attractive candidate for developing effective anti-sense and anti-gene reagents and drugs. PNAs have been found to inhibit RNA polymerase, human telomerase, HIV reverse transcriptase, and many more. Such PNAs are candidates for anti-cancer drugs and also as a means of developing novel drugs to treat HIV infections (AIDS). Despite these encouraging results, further progress is very much impeded by the inefficient uptake of PNA by living cells and the lack of efficient delivery systems.

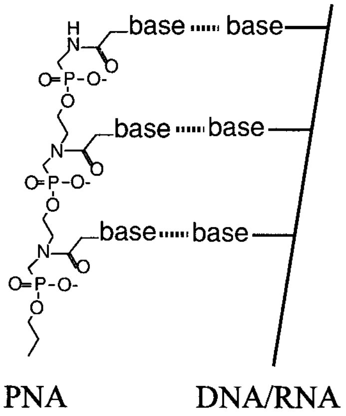


FIGURE 10 Structure of peptide nucleic acid (PNA). An artificial oligomer produced by chemical synthesis retains the ability to pair with bases, but is resistant to degradation by nucleases because its backbone does not contain the normal phosphodiester linkage.

SEE ALSO THE FOLLOWING ARTICLES

Bioconjugate Chemistry • DNA Testing in Forensic Science • Fiber-Optic Chemical Sensors • Gene Expression, Regulation of • Hybridomas, Genetic Engineering of • Ion Transport Across Biological Membranes • Protein Folding • Protein Structure • Protein Synthesis • Translation of RNA to Protein

BIBLIOGRAPHY

Eckstein, F. (2000). “Phosphorothioate oligodeoxynucleotides: What is their origin and what is unique about them?” *Antisense Nucl. Acid Drug Dev.* **10,** 117–121.

Efimov, V. A., Buryakova, A. A., and Chakhmakhcheva, O. G. (1999). “Synthesis of polyacrylamides *N*-substituted with PNA-like oligonucleotide mimics for molecular diagnostic applications,” *Nucl. Acids Res.* **27,** 4416–4426.

Kornberg, A., and Baker, T. A. (1992). “DNA Replication,” Freeman, New York.

Lewin, B. (2000). “Genes VII,” Oxford Univ. Press, New York.

Malik, S., and Roeder, R. G. (2000). “Transcriptional regulation through mediator-like coactivators in yeast and metazoan cells,” *Trends Biochem. Sci.* **25,** 277–283.

Mooney, R. A., Artsimovitch, I., and Landick, R. (1998). “Information processing by RNA polymerase: Recognition of regulatory signals during RNA chain elongation,” *J. Bacteriol.* **180,** 3265–3275.

Nielsen, P. E. (2000). “Peptide nucleic acids: On the road to new gene therapeutic drugs,” *Pharmacol. Toxicol.* **86,** 3–7.

Peterson, C. L., and Logie, C. (2000). “Recruitment of chromatin remodeling machines,” *J. Cell. Biochem.* **78,** 179–185.

Ray, A., and Norden, B. (2000). “Peptide nucleic acid (PNA): Its medical and biotechnical applications and promise for the future,” *FASEB J.* **14,** 1041–1060.

Stuart, K., Allen, T. E., Heidmann, S., and Seiwert, S. D. (1997). “RNA editing in kinetoplastid protozoa,” *Microbiol. Mol. Biol. Rev.* **61,** 105–120.

Sudarsanam, P., and Winston, F. (2000). “The Swi/Snf family—nucleosome-remodeling complexes and transcriptional control,” *Trends Genet.* **16,** 345–351.

Summers, D. K. (1996). “The Biology of Plasmids,” pp. 39–62, Blackwell Sci., Oxford.

Uhlmann, E. (1998). “Peptide nucleic acids (PNA) and PNA-DNA chimeras: From high binding affinity towards biological function,” *Biol. Chem.* **379,** 1045–1052.

Protein Folding

Maurice Eftink
Susan Pedigo
University of Mississippi

I. Introduction
II. Stability of the Tertiary Fold
III. Folding Pathways
IV. Empirical Approaches
V. Closing Comments

GLOSSARY

Absorbance spectroscopy Monitors conformational transitions in macromolecules by measuring absorbance changes, usually in the aromatic region of the ultraviolet (UV) spectrum.

Circular dichroism A very commonly used method for studying protein conformational changes.

Fluorescence The most sensitive of the commonly used optical methods for studying protein unfolding transitions.

Nuclear magnetic resonance A powerful method for studies with proteins, as there is such a large number of resolved signals.

Scanning calorimetry Measures the variation in the specific heat of a protein containing solution as a protein is thermally unfolded.

PROTEINS are only functional as enzymes, transport agents, receptors, and so forth when they exist in a folded, three-dimensional, native structure. The means by which this folded structure is achieved remains one of the most important questions in biochemistry and molecular biology. Current efforts toward sequencing the human genome and the genomes of other organisms have led to a large number of putative protein sequences. Much as the genetic code is the Rosetta stone that gave the link between DNA and protein sequences, we now need to find such a link between protein sequences and final structure and function. Without knowledge of the rules of protein folding, there can be little understanding of function from protein sequence information alone.

I. INTRODUCTION

We have interest in protein structure and function at both a fundamental and a practical level. There is astounding beauty in the mastery with which nature has tailored molecules for specific functions, activity levels, regulatory properties, and integration into complex macromolecular assemblies. As will be discussed, in most cases, these molecules assume a final stably folded structure

spontaneously. Thus, all of the information necessary for biological activity is contained in the simple sequence of amino acids as encoded by the DNA. Practically speaking, predicting protein structure, stability, and function from the primary sequence will open myriad opportunities in the areas of medicine (e.g., drug discovery and understanding molecular basis of disease), industry and manufacturing (e.g., biocatalysis and bioprocessing), and the environment (e.g., bioremediation).

Proteins are linear polymers of amino acids that are linked through amide linkages, commonly called the peptide bond. The "backbone" atoms include the amide linkages separated by a carbon that is derivatized by any one of 20 common side chains. The side chains may be grouped at neutral pH as acidic, basic, hydrophobic, and uncharged hydrophilic according to their chemical nature. Thus, although the backbone of the peptide polymer is a repeating identical unit, the side chains and their distinct properties dictate the nature of the protein. Because a subset of the amino acid side chains is charged at neutral pH (acidics are negative and basics are positive), the protein polymer is a polyelectrolyte. The linear sequence of amino acids is called the *primary structure* of the protein (Fig. 1). The primary structure dictates the way in which the polypeptide folds into a functional protein, in most cases without instructions from other sources.

Protein families are proteins related by structure or function. A protein family may be structurally diverse but have a particular cluster of amino acids at the active site that defines the class according to some catalytic function (e.g., dehydrogenases and kinases). Alternatively, proteins may have a structural motif that defines the class (e.g., helix–loop–helix motif of the EF-hand calcium-binding proteins). Proteins with identical function in different organisms often have slightly different primary structures (see below). The presence of certain amino acids relative to others in primary sequences allows putative protein sequences from the Human Genome Project, for example, to be classified into general protein families. Whether this initial classification is valid remains to be seen.

To discover the rules of protein folding, two major approaches have emerged: computational and empirical approaches. The computational approach, often termed *proteonomics*, attempts to predict the structure of a protein based on its sequence by defining a set of rules and criterion for their application. This topic is covered elsewhere in this series. The empirical approach to discovering the rules of protein folding defines global rules for folding based on lessons learned from particular proteins. These two methods are distinctly interwoven.[3] Hypotheses derived from one are testable through the other. In this paper, we will discuss the empirical approach to studying protein folding.

The empirical approach to understanding protein folding has relied heavily on mutational analysis. As mentioned earlier, proteins from different species with identical functions may have slightly different amino acid sequences, or mutations. Often the mutations are conservative, particularly in amino acids that are critical to the structure or function of the protein. Scientists study the different physical properties of these related proteins to gain insight into the role of amino acids in local or global structure and function of the protein. Often mutations are purposely engineered into protein sequences using molecular biological techniques to test hypotheses about roles of certain amino acids in structure or function. Selective substitution of tryptophan into a sequence allows placement of a convenient spectroscopic probe (see below).

Although proteins are very diverse, the one thing that almost all have in common is that they adopt *spontaneously* a unique and stable tertiary structure. This is an utter miracle of nature given the complexity of these heterogeneous polymers. The study of protein folding is focused on understanding the rules that govern the transition into and the stability of this unique fold. The transition into the tertiary structure is studied by kinetic methods. Thus, kinetic studies ask the question, "By what pathway is the final tertiary structure folded?" Alternatively, equilibrium thermodynamic methods ask "How stable is the final fold and why?" Each of these approaches will be discussed individually.

II. STABILITY OF THE TERTIARY FOLD

Stability of a protein is usually studied by observing the energetics of unfolding transitions given by the equations below:

$$N \leftrightarrow U \tag{1}$$

$$K_{un} = [U]/[N] \tag{2}$$

$$\Delta G^{\circ}_{un} = -RT \ln K_{un} \tag{3}$$

These equations apply to a simple two-state transition between the native (N) and the unfolded (U) state given by the equilibrium constant K_{un}. This is, by definition, a cooperative process without a detectable intermediate species. The denatured or unfolded state of a protein is generally considered to be an ensemble of conformations in which all parts of the protein are exposed to the solvent with a minimum of intramolecular interactions. The denatured state has high conformational entropy and is biologically inactive. The unfolding transition (Eq. (1) and Fig. 2) can be induced by pressure, temperature, extreme pH, and denaturants such as urea and guanidine

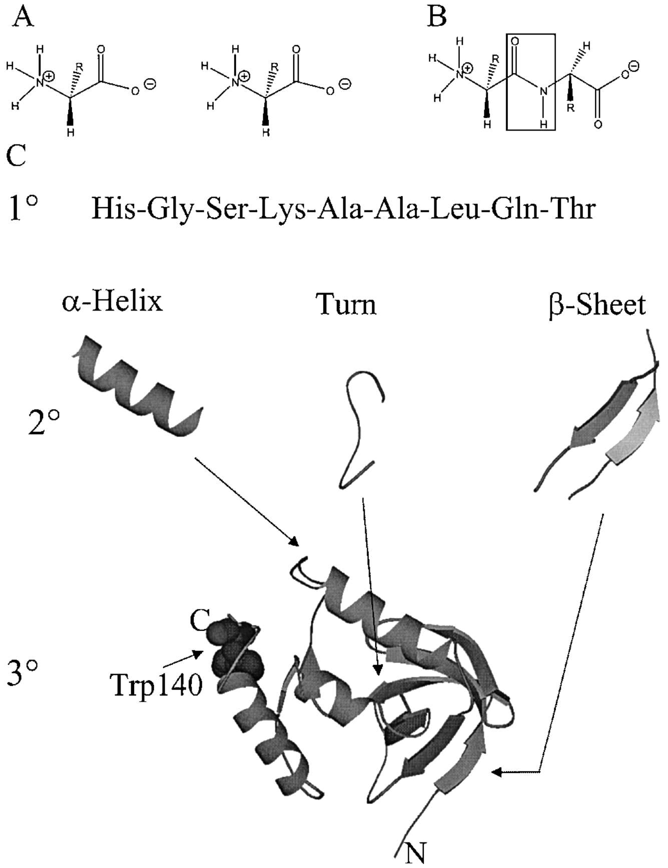


FIGURE 1 Diagram of the levels of protein structure. (A) Amino acids are the basic building blocks (monomers) of proteins (polymers). All amino acids contain a carboxylic acid group and an amine group connected by a central carbon called the α-carbon. Each of the 20 common amino acids, designated by a three-letter code, has a unique side chain (R) that is also bonded to the α-carbon. (B) Through a dehydrolysis reaction, an amide bond is formed (boxed region) that links the amino group of one amino acid to the carboxylic acid group of the next amino acid. (C) The primary structure of proteins (1°) is the linear sequence of amino acids written from the amino-terminal end (left) to the carboxy-terminal end (right), by convention. Secondary structure of proteins (2°) is classified into three major categories. In the α-helix structure, the backbone atoms (amide linkage and α-carbon) coil into a right-handed helical shape (residues 55 to 67 from staphylococcal nuclease[1] are shown). The α-helix is held together through a series of hydrogen bonds between the amide hydrogen and the carboxyl oxygen of the backbone atoms from amino acids further up the chain. The side chains (not shown) protrude from the central core structure like the spokes of a wheel. Another important secondary structure type is the turn (residues 76 to 88 from Staphyloccocal nuclease are shown). We use the term turn loosely here to represent the regions of proteins that turn corners, thereby allowing interactions between different and often distant (in terms of primary structure) substructures. The β-sheet structure is the third common secondary structure type (residues 9 to 12 and 72 to 76 from Staphylococcal nuclease are shown). It is similar to the α-helical structure in that hydrogen bonds between backbone atoms hold the structure together and the side chains (not shown) protrude from the structure above and below the plane of the sheet. In contrast to the α-helix, the β-sheet can be formed from segments of protein that are far apart in the primary sequence. The tertiary structure (3°) is the three-dimensional association of secondary structures into a unique and stable final fold. A ribbon tracing the backbone atoms of Staphylococcal nuclease is shown.[1,2]. The N-terminus of the protein is in the bottom right and the C-terminus is in the top left of the figure. No side chains are drawn except that of residues tryptophan 140.

HCl, as will discussed in a subsequent section. These perturbants disrupt the intramolecular interactions that hold proteins together. One can imagine that the ensemble of unfolded states could be influenced by the means used to unfold.

The native structure of proteins is stabilized by intramolecular, noncovalent interactions including hydrogen bonding, ionic, and van der Waals interactions, and covalent cross-links (disulfide bridges between cysteine residues) according to Eq. (4):

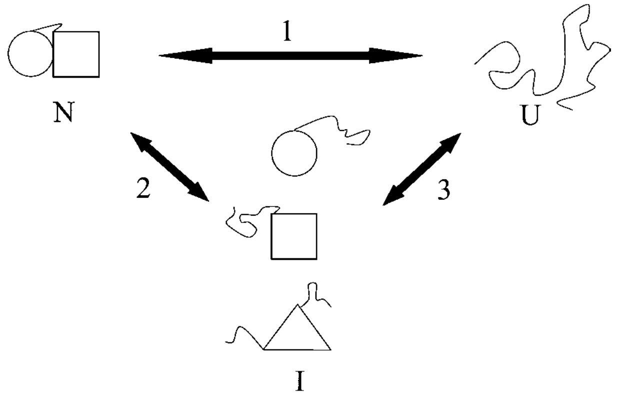


FIGURE 2 Illustration of cooperative vs. noncooperative unfolding transitions. If the native state of a protein (N) is denatured into the unfolded state (U) in a single transition (pathway 1), then it is a two-state or cooperative unfolding transition. Alternatively, the native state may be converted into one or more intermediate states (pathway 2). For example, if a protein is comprised of multiple domains, one of the domains may be unfolded first. It is also possible to form a completely different intermediate before unfolding completely. The presence of intermediate species may be observed using kinetic or equilibrium techniques. However, intermediates detectable by kinetic methods may or may not be observable by equilibrium methods.

$$\Delta G^{o}_{un} = \Delta G_{H\text{-}bond} + \Delta G_{ionic} + \Delta G_{vdW} + \Delta G_{S-S} + \Delta G_{H'phob} \quad (4)$$

Each term in Eq. (4) will be discussed separately. As mentioned earlier, an important stabilizing factor for the tertiary fold of a protein is its *intra*molecular hydrogen bonds ($\Delta G_{H\text{-}bond}$). Secondary structures are stabilized by hydrogen bonds between backbone amide atoms (Fig. 1). The side chains of neighboring secondary structural units can interact through hydrogen bonding. Ionic interactions (ΔG_{ionic}) between acidic and basic side chains may stabilize the tertiary structure of proteins and are pH dependent. The actual pKa of an ionizable side chain is influenced by the microenvironment in which it resides. Nonpolar and polar, but uncharged, amino acids interact through van der Waals interactions (ΔG_{vdW}). In some proteins, cysteine residues (side chain is a sulfhydryl) form disulfide linkages that can increase the overall stability of the protein (ΔG_{S-S}). Other possible factors not considered explicitly here are the effects of metals, nucleotides, prosthetic groups, and cofactors on protein structure and stability.

By far the most important noncovalent factor that determines protein stability is hydrophobic interactions ($\Delta G_{H'phob}$). In globular proteins, hydrophobic amino acids are buried in the interior where they create a "hydrophobic core." Although these nonpolar residues participate in van der Waals interactions, the primary driving force for the formation of the hydrophobic core is to avoid the aqueous solvent. Solvation of nonpolar side chains by aqueous solutions causes a decrease in the entropy of solution. To avoid this entropic penalty, proteins typically bury their nonpolar residues in the interior of a protein.[4]

III. FOLDING PATHWAYS

The intramolecular interactions discussed above stabilize the final folded structure of a protein. However, knowledge of the end states, N and U, tells us nothing of the path taken between them. Proteins fold on the time scale of microseconds to hundreds of seconds. It is impossible to sample all possible conformations during this time and it is clear that there is a preferred order of events leading to the final tertiary fold. Determining this order of events is an area of active inquiry. The questions that experimentalists are attempting to answer are "Do autonomously folding substructures nucleate the folding of other regions of the protein?" or "Do neighboring substructures fold and then collide to make the tertiary structure?" There is experimental evidence that hydrophobic amino acid residues collapse into a "hydrophobic core" and then the secondary structural units form around the core. It is likely that a combination of these scenarios leads to a correctly folded protein.

It is clear that the kinetics of protein folding is protein dependent. Some fold in a distinctly cooperative fashion, such that one can detect only the unfolded and native end states ($U \leftrightarrow N$), being two-state in a kinetic as well as equilibrium sense. This is equivalent to saying that there is a single rate-limiting step, and intermediate species are not populated. Alternatively, some proteins fold by populating one or more distinct intermediate species (e.g., $U \leftrightarrow I \leftrightarrow N$; see Fig. 2). Thus, formation of the intermediate species is fast, often formed in the dead-time of the instrument, and formation of the native species from the intermediate is relatively slow and easily monitored experimentally. It has been shown that this slow phase in some cases may be due to proline isomerization.[5]

IV. EMPIRICAL APPROACHES

A. General Experimental Strategies

As discussed above, experimental studies of protein folding reactions fall into the category of either equilibrium or kinetics studies, with the former yielding thermodynamic information about the energy differences between the native and denatured structural states and the latter studies providing information about the folding pathway and

the height of energy barriers between important species on this pathway. In general, to perform either an equilibrium (thermodynamics) or time-dependent (kinetics) study, one must be able to experimentally monitor a signal that tracks the population of the structural states of the protein.

There are a number of ways this can be done. The most convenient experimental methods involve solution-phase spectroscopic measurements; among these methods are absorption spectroscopy, fluorescence, circular dichroism, and nuclear magnetic resonance. Other methods include differential scanning calorimetry, light scattering, electrophoresis, and chromatography. This section gives a brief description of the advantages and disadvantages of some of the above methods. These methods are not equally applicable to equilibrium and time-dependent studies of protein unfolding, as some methods have a rapid response and some have a slow response. Methods also differ in their intrinsic sensitivity, which is related to the concentration of protein necessary to perform the measurement, their ease and economy of use, and whether they provide auxiliary information about the structure of the protein in its native and denatured states. What is meant by the last statement is that some of the spectroscopic signals can provide information about the secondary or tertiary structure of the protein species. For most types of spectroscopy, the signal arises from particular amino acid residues (e.g., aromatic side chains or peptide bond), thus differences in the signals for the conformational states can be related to differences in the local environment of these amino acid residues (e.g., tryptophan residue 140 in staphylococcal nuclease; see Fig. 1). If there are only a very few of such signal origination sites, then site-specific information can be obtained. If there are many probe sites and they are distributed throughout the protein's structure, then the method yields global information (e.g., signal from the amide linkage in the peptide backbone; see Fig. 1). It goes without saying that the protein sample to be studied must be well defined with regard to purity, and solution conditions must be selected and controlled to be relevant to other functional studies and studies with other proteins. Neutral pH, 20°C, and an ionic strength of 0.1 to 0.2 are the most commonly employed solution conditions.

A key to most of these methods and their use in protein unfolding studies is that the signal is a mole-fraction weighted average of the signals of each protein species. That is, for the simplest case of a thermodynamics study of the transition between a native, N, and unfolded, U, state of a protein, the observed signal, S, can be expressed as:

$$S = \sum S_i X_i \tag{5}$$

where X_i is the mole fraction of each species i and S_i is the intrinsic signal of species i. This relationship applies to most solution optical spectroscopic methods. Clearly, for a particular spectroscopic signal to be useful for tracking a $N \leftrightarrow U$ transition, the signal of the N and U states must be sufficiently different. The native (X_N) and unfolded (X_U) mole fractions are directly related to the equilibrium constant in Eq. (2), as:

$$X_N = 1/(1 + K_{un});\, X_U = K_{un}/(1 + K_{un}). \tag{6}$$

The transition from the native state to the unfolded state, or vice versa, can be induced in several ways, essentially by varying the solution conditions in a way that changes the equilibrium between the native and unfolded state. The transition may be induced by varying temperature, adding chemical (chaotropic agent) denaturant, adding acid or base, or increasing pressure. In the case of multimeric proteins, subunit dissociation, which may be accompanied by denaturation of the subunits, can be induced by dilution of the protein. Before discussing the various spectroscopic methods, some thermodynamic relationships are presented for describing the transitions induced in the above ways.

B. Basic Thermodynamic Relationships

Table I gives some widely accepted relationships for describing the variation of ΔG^{o}_{un} for a two-state $N \leftrightarrow U$ transition with temperature, chemical denaturant, pH, or pressure as the perturbations. One of the equations in Table I, when combined with those above and Eqs. (1–3), can be used to describe data as a function of the denaturing condition. The thermodynamic parameters related to the relationships in Table I are briefly described below.

1. *Thermal unfolding*: ΔH^{o}_{un} and ΔS^{o}_{un} are the enthalpy and entropy changes for a two-state unfolding reaction. Both ΔH^{o}_{un} and ΔS^{o}_{un} may be temperature dependent, when the heat capacity change, ΔC_p, has a nonzero value. In this case, Eq. (7b) in Table I (the Gibbs-Helmholtz equation) should be used, where the $\Delta H^{o}_{o,un}$ and $\Delta S^{o}_{o,un}$ are values at some defined reference temperature, T_o (e.g., 0° or 20°C).[6,7] The heat capacity change for unfolding of proteins is typically found to be positive and to be related to the increase in solvent exposure of apolar side chains upon unfolding. That is, a positive ΔC_p is a result of the hydrophobic effect. A consequence is that the $\Delta G^{o}_{un}(T)$ for unfolding of a protein will have a parabolic dependence on temperature and will show both high-temperature and low-temperature induced unfolding.[8]

2. *Denaturant-induced unfolding*: The empirical relationship in Table I for chemical denaturation includes

TABLE I Relationships Describing Two-State Transitions in Proteins

Temperature

$$\Delta G_{un}(T) = \Delta H^{o}_{un} - T\Delta S^{o}_{un} \quad (7a)$$

$$\Delta G_{un}(T) = \Delta H^{o}_{o,un} + \Delta C_p(T - T_o) - T[\Delta S^{o}_{o,un} + \Delta C_p \ln(T/T_o)] \quad (7b)$$

where

$\Delta H^{o}_{o,un}$ is the enthalpy change at $T = T_o$.

ΔS^{o}_{un} is the entropy change at $T = T_o$.

ΔC_p is the change in heat capacity upon unfolding.

Chemical Denaturants

$$\Delta G_{un}([\mathbf{d}]) = \Delta G^{o}_{o,un} - m[\mathbf{d}] \quad (\textit{linear extrapolation model}) \quad (8)$$

where

$\Delta G^{o}_{o,un}$ is the free energy change in the absence of **d**.

$m = \delta\Delta G_{un}/\delta[\mathbf{d}]$.

pH

$$\Delta G_{un}(pH) = \Delta G^{o}_{o,un} - RT \bullet \ln\left\{\frac{\left(1 + \frac{[H^+]}{K_{a,U}}\right)^n}{\left(1 + \frac{[H^+]}{K_{a,N}}\right)^n}\right\} \quad (9)$$

where

$\Delta G^{o}_{o,un}$ is the free energy change at neutral pH.

$K_{a,U}$ is the acid dissociation constant of a residue in the unfolded state.

$K_{a,N}$ is the acid dissociation constant of a residue in the native state.

Pressure

$$\Delta G_{un}(P) = \Delta G^{o}_{o,un} - \Delta V_{un}(P_o - P) \quad (10)$$

where

$\Delta V_{un} =$ volume change for $N \leftrightarrow U$ transition.

$P_o =$ reference pressure.

For a two-state transition, $A \leftrightarrow B$ (or $N \leftrightarrow U$ for the unfolding of a native, N, to an unfolded, U, state of a protein) the mole fractions of the N and U states are given as $X_N = 1/Q$, $X_U = \exp(-\Delta G_{un}/RT)/Q$, where $Q = 1 + \exp(-\Delta G_{un}/RT)$ and the function for ΔG_{un} is taken from above the average fluorescence signal, $F_{calc} = \sum X_i(F_i + x\delta F_i/\delta_x)$, where x is a generalized perturbant.

$\Delta G^{o}_{o,un}$, the free energy change for unfolding in the absence of denaturant, and m, the denaturant susceptibility parameter ($= -\delta\Delta G_{un}/\delta[d]$), where $[d]$ is the molar concentration of added chemical denaturant.[9,10] Through an empirical relationship, the given equation appears to adequately describe the pattern for denaturant-induced unfolding of a number of proteins. The $\Delta G^{o}_{o,un}$ value is a direct measure of the stability of a protein at the ambient solvent conditions, which can be moderate temperature and pH (e.g., 20°C and pH 7). The m value also provides structural insights, as m values have been suggested to correlate with the change in solvent accessible apolar surface area upon unfolding of a protein.[11] For example, a relatively large m value (i.e., a high susceptibility of the unfolding reaction to denaturant concentration) indicates that there is a large change in the exposure of apolar side chains on unfolding, which might be the case for a protein that has an extensive core of apolar side chains that are exposed upon denaturation.

3. *Acid-induced unfolding*: The relationship for acid-induced unfolding assumes that there are n equivalent acid dissociating groups on a protein that all have the same $pK_{a,U}$ in the unfolded state and that they are all perturbed to have a $pK_{a,N}$ in the N state. If the $pK_{a,N}$ is more than 2 pH units lower than $pK_{a,U}$, then the equation simplifies with the denominator of the right term going to unity. The simplest relationship for acid-induced unfolding includes $\Delta G^{o}_{o,un}$, the free energy of unfolding at neutral pH; n, the number of perturbed acid dissociating residues; and their $pK_{a,U}$ in the unfolded state. Presumably, n should be an integer and $pK_{a,U}$ should be approximately equal to the values for such amino acids as glutamate, aspartate (e.g., $pK_{a,U}$ should be about 4 to 4.3) or histidine (e.g., $pK_{a,U}$ should be around 6.5).

4. *Pressure-induced unfolding*: In the relationship for pressure, P, induced unfolding of proteins, $\Delta G^{o}_{o,un}$ is again the value of the free energy change at 1 atmosphere pressure and $\Delta V_{un} = V_U - V_N$ is the difference in volume

of the unfolded and native states. Pressure-induced unfolding studies require a specialized high pressure cell.[12,13]

5. *Dissociation/unfolding of oligomeric proteins*: Oligomeric proteins are interesting as models for understanding intermolecular protein-protein interactions. A general question for oligomeric proteins, including the simplest dimeric (D) proteins, is whether the protein unfolds in a two-state manner, $D \leftrightarrow 2U$, or whether there is an intermediate state, which might be either an altered dimeric state, D', or a folded (or partially folded) monomer species, M. Models for these two situations are as follows:

$$D \leftrightarrow D' \leftrightarrow 2U \tag{11a}$$

$$D \leftrightarrow 2M \leftrightarrow 2U \tag{11b}$$

For a $D \leftrightarrow 2U$ model, the relationships between the observed spectroscopic signal, S_{exp}; the mole fraction of dimer, X_D, and unfolded monomer, X_U; and the unfolding equilibrium constant ($K_{un} = [U]^2/[D]$) will be given by Eq. (5) and

$$X_U = \frac{K_{un}^2 + 8K_{un}[P]_0^{1/2} - K_{un}}{4[P]_0}; X_D = 1 - X_U \tag{12}$$

where $[P]_0$ is the total protein concentration (expressed as monomeric form), where S_i is the relative signal of species i and where K_{un} will depend on the perturbant as given by one of the above equations. That is, the transition should depend on the total subunit concentration, $[P]_0$, and on any other perturbation axis.

C. Experimental Signals

1. Absorbance Spectroscopy

Absorbance spectroscopy (difference spectroscopy) monitors conformational transitions in macromolecules by measuring absorbance changes, usually in the aromatic region of the ultraviolet (UV) spectrum. The amino acids tryptophan and tyrosine are the most important chromophores in the UV region for proteins. As mentioned earlier, tryptophan residues are often engineered into proteins as reporters of local and/or global environment.

The indole ring of tryptophan and the phenol ring of tyrosine show sensitivity of their absorbance spectrum to solvent polarity. There is a blue shift in the absorbance of indole and phenol upon increasing solvent polarity. As a result, there will often be a blue shift in the absorbance of tryptophan (typically monitored as a decrease in absorbance in the 291- to 294-nm region of the spectrum) or tyrosine (at 285 to 288 nm) upon unfolding of a protein and a consequent increase in the exposure of these aromatic side chains to water.[14] Tryptophan's absorbance is also sensitive to the local electrostatic field; changes in indole-charge interactions can cause either red or blue shifts upon protein unfolding.[15]

Table II gives the typical concentration range used for unfolding studies with proteins using this and other methods. The sensitivity of difference absorbance measurements will depend on the molar extinction coefficient of the chromophore and their number, but a concentration range of 0.01 to 0.1 m*M* protein is usually needed for reasonable signal to noise with a 1-cm pathlength cell. Thermal scans, to induce the unfolding transition, are easy to perform with accessories available for most absorbance spectrophotometers. Chemical denaturant- or pH-induced transitions can be less convenient (unless one has automated titration equipment), since a series of solutions with equal protein concentration and varying denaturant must be prepared. With any of these perturbing conditions, it is important to realize that the variation in the conditions itself (i.e., varying temperature, pH, chemical composition) can lead to a "baseline" change in the absorbance signal from the native and unfolded species.[16] So long as these baseline trends are linear and not as large as the absorbance change associated with the conformational transition, the baseline trends can be corrected for in the data analysis.

The advantages of absorbance measurements are the ready availability, ease of use, and low cost of the instrumentation. The biggest disadvantage is that it is less sensitive than some other methods.

2. Circular Dichroism

Circular dichroism (CD) is a very commonly used method for studying protein conformational changes. The far UV spectral region (180 to 250 nm) is dominated by absorbance by peptide bonds, and there are signature spectra for α-helix and other types of secondary structure in a protein. Additionally, the aromatic CD spectral region of 250 to 300 nm senses the chirality around the aromatic amino acid side chains and there is usually a structured aromatic CD spectrum for the native state of a protein.[14,17,18]

The effective sensitivity of CD is comparable to or slightly better than that of difference UV absorbance spectroscopy. CD instruments can be purchased with thermoelectric cell holders for thermal scans and with automated titrator syringe pumps for chemical denaturant titrations. Since the far-UV spectral regions is important in protein unfolding studies, it is necessary to work with salts and buffers that have minimal absorbance in this region. When performing CD measurements, it is necessary to pay attention to the buffer and salts and other solution components (e.g., chemical denaturants) being

TABLE II Solution Methods for Monitoring the Progress of Protein Unfolding Transitions

Method	Conc. Range (m*M*)[a]	Scanning or Titrations[d]	Structure Sensed	Kinetic Applications
Absorbance	0.01–1	TS/AT	Local	***
Circular dichroism	0.01–0.1	TS/AT	Secondary	***
Fluorescence	0.0001–0.01	TS/AT	Local/tertiary	***
FTIR	0.5–2	TS	Secondary	*
Light scattering	0.1–1	No	Size and shape	*
NMR	1–10	No	Local/tertiary	*
DSC	0.02–0.2	TS	Tertiary	—[e]
Activity/binding	—[b]	P	Tertiary	*
Chemical reactivity	variable	P	Local/tertiary	*
Chromatography	—[c]	No	Size and shape	—
Electrophoresis	—[c]	Gradients	Size, shape, charge	—
Potentiometry	0.1–1	No	Local	—

[a] Concentration ranges are for typical experiments with a 20-kDa protein.

[b] The concentration range will depend on the method being used to measure enzymatic activity or ligand binding.

[c] The concentration of protein varies during the course of the experiment as the sample flows through the column, gel, or capillary. Initial concentrations are usually in the range of 1 mg/mL.

[d] "TS" refers to the ability to perform thermal scans to unfold a protein; "AT" refers to the ability to perform automated titrations of a protein sample with chemical denaturant, acid, or base while the sample is loaded in the instrument. The label "P" indicates that an automated thermal scan or titration may be possible for certain applications, though this is not commonly done. The "Structure Sensed" column lists the features of the protein structure (e.g., secondary and tertiary structure, local interactions, etc.) that are sensed by the method. Some of these entries are judgment calls. The "Kinetic Applications" column indicates the amenability of the method to protein folding/unfolding kinetics experiments. A label "***" indicates that transient mixing or other means are available for the rapid initiation of the reaction. A label "*" indicates that the method is amenable to study relatively slow reactions (i.e., by a hand-mixing experiment).

[e] Through variation of thermal scan rate or a frequency domain application of DSC, it is possible to obtain kinetics information.

used, particularly if one wishes to make measurements below 200 nm, as various buffers, salts, and denaturants can absorb a significant amount of light in the far-UV. Schmid[14] has provided a number of practical tips regarding the application of CD for studies with proteins. There is less interference by buffer, salts, etc. in the aromatic UV spectral region. Whereas the aromatic CD signals can sense the loss of tertiary structure in a protein as it denatures, the CD signals in this region are much smaller than those in the far-UV CD region, giving a lower signal-to-noise ratio. Baseline slopes, as one varies temperature or chemical denaturant, also must be considered in CD measurements in both the far-UV and aromatic spectral region; however, the baselines trends are usually not large.

A difference between far-UV CD and other optical methods is that CD signals observe changes throughout the structure of the protein (i.e., its secondary structure) and the magnitude and direction of the signal changes can be more directly related to changes in structure (e.g., a loss of ellipticity at 222 nm can be related to a loss of α-helix).

3. Fluorescence

Fluorescence is the most sensitive of the commonly used optical methods for studying protein unfolding transitions.[14,19–21] The absolute sensitivity depends on a number of factors (e.g., lamp or laser intensity, cell pathlength, chromophore extinction coefficient, and quantum yield), of course, but commercial fluorometers can usually detect signals down to the 10-n*M* range. Either intrinsic or extrinsic fluorophores can be used. The most commonly used intrinsic fluorophores are the tryptophan and tyrosine residues, with the former being the most important due to its larger molar extinction coefficient and a redder absorbance and emission. The fluorescence of tryptophan residues is very dependent on the local microenvironment of its indole side chain, making tryptophan fluorescence responsive to the structure

of a protein. This spectral responsiveness is in terms of its emission maximum and its quantum yield. For example, the emission maximum of tryptophan almost always shifts to longer wavelengths (red shifts) upon unfolding a protein and increasing the solvent exposure of this amino acid side chain. There is a large literature about the fluorescence of tryptophan residues in proteins and its use to study changes in the structure of proteins.[19]

A variety of extrinsic fluorophores can be attached to proteins to serve as fluorescence probes. These can be selected to maximize sensitivity and to avoid contamination (i.e., by moving to longer absorption and emission wavelengths) from other absorbing components.[22] With both intrinsic and extrinsic fluorescence probes, the method focuses only on these probes sites, which might be as few as a single site on a protein.

Like the signals from absorption spectroscopy and CD, fluorescence intensity signals (either at a single wavelength or integrated over the emission envelope) follows Eq. (5) and can be used to extract thermodynamic information. However, there are other easily measured fluorescence signals (emission maximum and anisotropy) that do not follow the mole fraction averaging of Eq. (5).[19] The apparent emission maximum of a protein will be dominated by the structural state, native or unfolded, which has the higher quantum yield. Consequently, the apparent emission maximum will frequently not give a true reflection of the population of native and unfolded states, thus limiting the value of this type of fluorescence measurement for use in recovering thermodynamic parameters. (Rather than use the apparent emission maximum, it is better to perform curve fitting with composite spectra of the native and unfolded states.)

Fluorescence anisotropy values for the fluorescence of a fluorophore on a protein will depend on the fluorophore's rotational freedom and fluorescence lifetime. Because the motional freedom of intrinsic or extrinsic fluorophores will usually increase when a protein unfolds, a change in a protein's fluorescence anisotropy is expected upon unfolding. However, to properly use anisotropy to analyze the thermodynamics (or kinetics) of an unfolding transition, Eq. (1) should be replaced with one that includes the fluorescence quantum yield of the protein's structural states (see Reference 19).

As with the above-listed optical methods, fluorescence instruments are designed to allow automated thermal scans and/or titrations. The baseline problem can be more significant with fluorescence than the other methods and should not be ignored. In particular, it is well known that the fluorescence intensity of fluorophores will decrease with increasing temperature, regardless of whether there is a conformational transition. While baseline trends may not

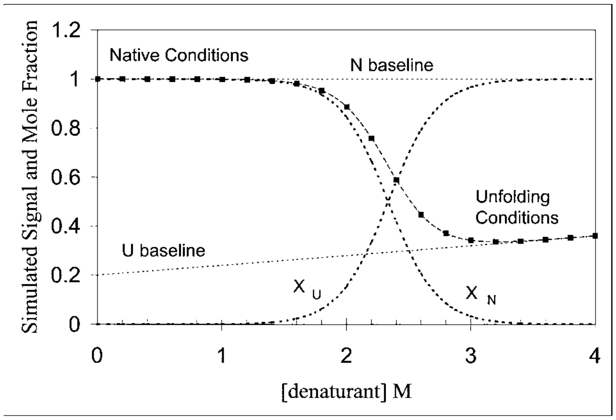


FIGURE 3 Simulation of denaturant-induced unfolding of a protein in a two-state manner. A simulated fluorescence signal (■) is plotted vs. denaturant concentration for a protein, using Eq. (8). The simulated fluorescence signal decreases with addition of denaturant because the unfolded species has a smaller fluorescence signal (fluorescence is quenched on unfolding). The pre- and post-transition baselines may slope, as shown. The fraction of unfolded species (X_U) increases from left to right as the fraction of native species (X_N) decreases.

be linear over extensive ranges of the perturbing variable (e.g., temperature or chemical denaturant), it is usually adequate to assume linear slopes over a limited range of the variable.

The advantages of fluorescence for studying protein unfolding reactions are the wide concentration range that can be measured and the responsiveness of the signal to the microenvironment of the fluorophore. Additionally, fluorescence signals of the native and unfolded state can provide a modicum of structural information about these states (at least with respect to the microenvironment of the fluorophores). Figure 3 shows simulated data for the denaturant-induced unfolding of a protein, as would be monitored by fluorescence intensity measurements.

4. Differential Scanning Calorimetry

Another frequently used method is differential scanning calorimetry (DSC), which measures the variation in the specific heat of a protein containing solution as a protein is thermally unfolded.[23–25] As opposed to the above optical techniques, where photons are being measured, calorimetry measures the transfer of heat associated with the thermally induced conformational transition. DSC and related types of calorimetry are intrinsically less sensitive than the optical methods. Nonetheless, advances in the technique have made it possible to perform DSC studies with samples as low as 0.1 mg/mL.

Temperature is scanned in DSC measurements, so it is the variable that causes the structural transition of a protein. DSC data are typically presented as thermograms that

yield a heat-capacity maximum corresponding to the thermal transition temperature, T_G, and an enthalpy change, ΔH_{un}, for the transition. The ΔH_{un} value can be determined either by integration of the thermogram or by curve fitting (i.e., fitting a van't Hoff equation to the shape of the thermogram). Referring to these two ΔH_{un} estimates as the calorimetric and van't Hoff values, the ratio of the calorimetric and van't Hoff ΔH_{un} values can be used to determine whether the transition is best described as a two-state process. That is, a ratio of 1.0 (indicating that two ΔH_{un} estimates are essentially the same) means that the structural transition is two state.

5. Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) spectroscopy is a powerful method for studies with proteins, as there is such a large number of resolved signals (due to the individual nuclei, such as the ^{1}H and ^{13}C atoms in the backbone and/or side chains of the amino acids).[26,27] This gives the potential to track conformational transitions by observing changes at a large number of individual sites on the protein. This is further made possible by the fact that the signals (peaks having various chemical shifts) are usually widely dispersed in the native state of a protein, as a consequence of the sensitivity of the resonance peak for individual nuclei to the local magnetic field, which in turn is related to the three-dimensional structure of the protein. Unfolded proteins, by comparison, usually have a much narrower range of resonance peaks for similar amino acid components.

Tracking any of the individual resonance signals, such as those assigned to histidine or tryptophan residues, as a function of denaturing condition (e.g., temperature, pH, or added chemical denaturant) provides a way to study the unfolding process, as the signal is transformed from that of the native state to that of the unfolded state. An important difference between NMR and the above optical methods is that NMR signals can be dynamically averaged signals or individual signals can appear for the native and unfolded states. The latter results if the rate of interconversion of the conformational states is relatively slow in comparison to the difference in resonance frequencies of the signals for the two states.

The use of NMR for proteins studies is usually limited to proteins having molecular weight of about 25 kDa or less. The method requires a relatively high concentration of protein, compared to other methods.

Besides the above application of NMR to track the population of native and unfolded states, NMR also can provide very high-quality information about the tertiary and secondary structure of proteins. In addition, pulsed isotope labeling experiments can provide information about the pathway for protein folding reactions and can provide estimates of the unfolding equilibrium constant at individual sites on the protein.[26,27]

6. Other Experimental Methods

Fourier transform infrared (FTIR) vibrational spectroscopy senses the hydrogen bonding pattern of the peptide bonds of a protein and can detect unfolding transitions in terms of changes in the secondary structure patterns.[28] As compared to CD, which also senses secondary structure, FTIR is relatively more responsive to β-sheet structures. A disadvantage of FTIR is that it requires a higher protein concentration and that it is more difficult to automate for titration experiments.

Light-scattering methods, such as small-angle X-ray scattering, or quasi-elastic light scattering, can provide information about the size of a protein, in terms of its radius of gyration. Unfolding or aggregation reactions are detected as increases in the hydrodynamic radius.[29] These scattering methods are also relatively difficult to adapt to temperature or titration experiments.

The ability of a protein to bind a specific ligand or to have catalytic activity can be used to determine the population of native species. The possibilities are numerous, depending on the way that activity and binding are measured. These activity/binding assays should be easy to automate for a series of denaturant concentrations or pH values.

Size exclusion chromatography and gel or capillary electrophoresis are methods that separate protein molecules based on size (or size and charge).[30–32] In these methods, the protein sample travels down the column, gel slab, or capillary and, for a pure protein, should exit as a single peak traveling past the detector. Denatured proteins should appear to have a larger hydrodynamic radius and should travel more slowly. If the kinetics of interconversion of the native and unfolded species is slower than the time needed to travel through the column (gel or capillary), then it is possible to detect individual peaks for the native and unfolded species. If the interconversion is rapid, a kinetically averaged peak position will be observed.

An example of a potentiometric measurement is one in which the pH (or number of protons bound versus those bound at some reference condition) is measured as a function of the denaturing condition. Such an approach would require a difference in the pK_a of one or more amino acid side chains in the native and unfolded state. Usually, several such amino acid side chains are in a protein. However, the potentiometric approach requires technical skill, and it is difficult to use in combination with high concentrations of chemical denaturants or temperatures far from ambient.

D. Kinetics Experiments

Studying the kinetics of folding or unfolding a protein involves the rapid initiation of the reaction by either removal of the denaturing condition (to initiate folding) or addition of the denaturing condition (unfolding). The most common way in which this is done is by a stopped-flow mixing device, in which a solution of the protein in one solution condition (e.g., neutral buffer with no chemical denaturant) is rapidly mixed with a second solution (e.g., concentrated chemical denaturant or acid). Temperature or pressure can also be used as the perturbing condition. Upward temperature jumps can be initiated by high-powered laser pulses or electrical discharges in the solution. Downward pressure jumps can be initiated by releasing some type of valve after high pressure has been established. Other ways to rapidly initiate a protein folding or unfolding reaction include such things as laser flash-induced chemical reactions, which can dissociate heme-carbon monoxide bonds of heme proteins. The remaining discussion will emphasize stopped-flow mixing reactions, since these are the most widely available approaches.

For small, globular proteins, the kinetics of denaturant-induced or acid-induced unfolding reactions are often found to be described as a mono-exponential process, indicating that there is a single energy barrier between the native and unfolded species. For larger proteins, particularly those with multiple domain structures, one often finds the kinetics to be more complicated. In those cases, the folding or unfolding reaction may be described by more than one exponential decay term. This can be an indication of the existence of multiple, slowly interconverting unfolded states, the existence of multiple energy barriers along the folding/unfolding pathway, the existence of more than one pathway, or the existence of some off-path (or dead-end) species. The challenge in kinetics experiments is to first determine the minimum number of decay terms needed to describe the reaction and to then determine a reaction mechanism consistent with the kinetics data. Often, unique mechanisms cannot be determined and it is the art of the scientist to establish which mechanism is most reasonable for the system being studied. In many cases research will focus on determining the number of intermediates on the folding pathway and in trying to gain structural information about these intermediates.

Whether studying folding or unfolding, the reaction should be described with the following general empirical relationship:[33,34]

$$S(t)_{[d]} = S_{\infty,[d]} + \sum_i \Delta S_{i,[d]} \cdot \exp(-t/\tau_i) \qquad (13)$$

where $S(t)_{[d]}$ is a time-dependent change in some generalized signal at denaturing condition described by $[d]$, $S_{\infty,[d]}$, is the signal at equilibrium (infinite time after initiation of reaction) at the denaturing condition, and $\Delta S_{i,[d]}$ is the signal amplitude associated with relaxation time constant, τ_i. If there is only a single step in the process ($i = 1$), the reaction will be a mono-exponential. With more steps, the reaction becomes bi-, tri-, . . . exponential, and fitting the above equation to the data can be difficult. Even with the number of terms and the corresponding values of τ_i determined, it is still a challenge to relate the τ_i and the associated amplitudes to a reaction mechanism. How this can be done is beyond the scope of this entry and we refer readers to References 35 and 36.

To experimentally apply the above equation, it is necessary to monitor some signal that responds to the structural state of the protein, just as is the case in equilibrium studies. For very rapid folding/unfolding reactions, the monitoring method must itself be able to respond more rapidly (with adequate signal to noise) than the chemical reaction. Fluorescence, absorbance, and circular dichroism are fairly rapidly responding optical methods and enjoy widespread use in combination with stopped-flow mixers for this purpose. As indicated in Table II, some of the other methods are not easily adapted for the rapid initiation and continuous monitoring of a reaction's progress. Obviously, when a reaction occurs on the time scale of minutes to hours, then a hand-mixing experiment can suffice.

V. CLOSING COMMENTS

The empirical approaches mentioned above have afforded great insight into the transition from a linear sequence of amino acids into a final tertiary structure. Researchers in the field continue to pioneer new techniques that give insight into the complex inter- and intramolecular interactions that dictate structure and function of proteins. These efforts coupled with computational approaches are helping to reveal the rules that nature uses to create the complex and unique structure of proteins.

SEE ALSO THE FOLLOWING ARTICLES

BIOPOLYMERS • CHEMICAL THERMODYNAMICS • ELECTROPHORESIS • NUCLEAR MAGNETIC RESONANCE • PROTEIN STRUCTURE • TRANSLATION OF RNA TO PROTEIN

BIBLIOGRAPHY

Hynes, T. R., and Fox, R. O. (1991). "The crystal structure of staphylococcal nuclease refined at 1.7 A resolution," *Proteins Struct. Funct. Genet.* **10,** 92–105.

Kraulis , P. J. (1991). "MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures," *J. Appl. Crystallogr.* **24,** 946–950.

Laurents, D. V., and Baldwin, R. L. (1998). "Protein Folding: Matching Theory and Experiment" *Biophysical J.* **75,** 428–434.

Privalov, P. L., and Gill, S. J. (1988). "Stability of protein structure and hydrophobic interaction," *Adv. Protein Chem.* **39,** 191–234.

Lin, L. N., and Brandts, J. F. (1988). "Separation of the nativelike intermediate from unfolded forms during refolding of ribonuclease A," *Biochemistry* **27,** 9037–9042.

Schellman, J. A. (1987). "The thermodynamic stability of proteins," *Ann. Rev. Biophys. Biophys. Chem.* **16,** 115–137.

Becktel, W. J., and Schellman, J. A. (1987). "Protein stability curves," *Biopolymers* **26,** 1859–1876.

Privalov, P. L. (1990). "Cold denaturation of proteins," *Crit. Rev. Biochem.* **25,** 281–305.

Schellman, J. A. (1978). "Solvent denaturation," *Biopolymers* **17,** 1305–1322.

Pace, C. N. (1986). "Determination and analysis of urea and guanidine hydrochloride denaturation curves," *Methods Enzymol.* **131,** 266–280.

Santoro, M. M., and Bolen, W. D. (1992). "Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants," *Biochemistry* **31,** 8063–8068.

Palidini, A. A., and Weber, G. (1981). "Pressure induced reversible dissociation of enolase," *Biochemistry* **20,** 2587–2593.

Silva, J. L., and Weber, G. (1993). "Pressure stability of proteins," *Annu. Rev. Phys. Chem.* **44,** 89–113.

Schmid, F. X. (1989). "Spectral methods of characterizing protein conformation and conformational changes," *In* "Protein Structure: A Practical Approach" (T. E. Creighton, ed.), pp. 251–285, IRL Press, Oxford.

Andrews, L. J., and Forster, L. S. (1972). "Protein difference spectra. Effect of solvent and charge on tryptophan," *Biochemistry* **11,** 1875–1879.

Eftink, M. R. (1995). "Use of multiple spectroscopic methods to monitor the equilibrium unfolding of proteins," *Methods Enzymol.* **259,** 487–512.

Johnson, W. C. (1988). "Secondary structure of proteins through circular dichroism spectroscopy," *Ann. Rev. Biophys. Biophys. Chem.* **17,** 145–166.

Yu, S., Venyaminiov, S. Y., and Yang, J. T. (1996). "Determination of protein secondary structure," *In* "Circular Dichroism and the Conformational Analysis of Biomolecules" (G. D. Fasman, ed.), pp. 69–107, Plenum Press, New York.

Eftink, M. R. (1996). "Fluorescence techniques for studying protein structure," *Methods Biochem. Anal.* **35,** 127–205.

Eftink, M. R. (1994). "The use of fluorescence to monitor unfolding transitions in proteins," *Biophys. J.* **66,** 482–501.

Lakowicz, J. R. (200). "Protein fluorescence," *In* "Topics in Fluorescence Spectroscopy," Vol. 6, Plenum Press, New York.

Haugland, R. P. (1983). "Covalent fluorescent probes," *In* "Excited States of Biopolymers" (R. F. Steiner, ed.) pp. 29–58, Plenum Press, New York.

Sturtevant., J. M. (1897). "Biochemical applications of differential scanning calorimetry," *Ann. Rev. Phys. Chem.* **38,** 463–488.

Biltonen, R. L., and Freire, E. (1978). "Thermodynamic characterization of conformational states of biological macromolecules using differential scanning calorimetry," *CRC Critical Rev. Biochem.* **5,** 85–124.

Cooper, A., and Johnson, C. M. (1994). "Differential scanning microcalorimetry," *Methods Molec. Biol.* **22,** 125–136.

Roder, H. (1989). "Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange," *Methods Enzymol.* **176,** 446–473.

Englander, W., and Mayne, L. (1992). "Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR," *Ann. Rev. Biochem. Biophys. Chem.* **21,** 243–265.

Muga, A., Mantsch, H. H., and Surewicz, W. K. (1991). "Membrane binding induces destabilization of cytochrome c structure," *Biochemistry* **30,** 7219–7224.

Nicoli, D. F., and Benedek, G. B. (1976). "Study of thermal denaturation of lysozyme and other glubular proteins by light-scattering spectroscopy," *Biopolymers* **15,** 2421–2437.

Corbett, R. T., and Roche, R. S. (1984). "Use of high-speed size-exclusion chromatography for the study of protein folding and stability," *Biochemistry* **23,** 1888–1894.

Creighton, T. E. (1980). "Kinetic study of protein unfolding and refolding using urea gradient electrophoresis," *J. Mol. Biol.* **137,** 61–80.

Hilser, V. J., Worosila, G. D., and Freire, E. (1993). "Analysis of thermally induced protein folding/unfolding transitions using free solution capillary electrophoresis," *Anal. Biochem.* **208,** 125–131.

Dandliker, W. B., Dandliker, J., Levison, S. A., Kelley, R. J., Hicks, A. N., and White, J. U. (1978). "Fluorescence methods for measuring reaction equilibria and kinetics," *Methods Enzymol.* **48,** 380–415.

Eftink, M. R., and Shastry, M. R. C. (1997). "Fluorescence methods for studying kinetics of protein folding reactions," *Methods Enzymol.* **278,** 258–287.

Utiyama, H., and Baldwin, R. L. (1986). "Kinetic mechanisms of protein folding," *Methods Enzymol.* **131,** 51–70.

Schmid, F. X. (1992). "Kinetics of unfolding and refolding of single-domain proteins," *In* "Protein Folding" (T. E. Creighton, ed.), pp. 197–241, W. H. Freeman, New York.

Protein Structure

Ivan Rayment
University of Wisconsin, Madison

I. Introduction
II. Amino Acids
III. Protein Structure Determination
IV. Structural Hierachy
V. Secondary Structural Motifs
VI. Protein Stability
VII. Tertiary Structure
VIII. Membrane Proteins
IX. Superfamilies and Structural Evolution
X. Quaternary Structure
XI. Conclusions

GLOSSARY

Primary structure Linear sequence of amino acids in a polypeptide.

Quaternary structure Arrangement of polypeptides in macromolecular assembly.

Secondary structure A description of the three-dimensional structure adopted by a localized sections of the polypeptide chain.

Tertiary structure A description of the arrangement of secondary structural elements within the protein.

Units of length 1 Å = 0.1 nm.

PROTEINS are linear polymers of amino acids linked by amide bonds where the biological function is dictated by the sequence of amino acid residues within the polymer. Most proteins adopt a well-defined three-dimensional structure to fulfil their biological role. Thus, knowledge of protein structure is central to understanding the molecular basis of life.

I. INTRODUCTION

Proteins mediate the majority of biological processes. All proteins share the common feature that they are condensation polymers of amino acids whose sequence is specified by the genetic information contained within the genome of the organism. Complete DNA sequences for organisms ranging from *Escherichia coli* to humans suggest that the total number of proteins necessary for life lies in the range of 4200–50,000, although the number of genes in higher organisms is still under debate. Most of these proteins

adopt a well-defined three-dimensional structure in solution that is essential for protein function. Indeed unfolding or denaturation of a protein typically leads to a loss of biological activity.

The amino acid sequence of a protein contains all of the information necessary to dictate its final three-dimensional structure or fold. In many cases small proteins can be unfolded and refolded *in vitro* without loss of activity. In more complex proteins, chaparones are frequently necessary to allow a protein to reach its properly folded or correct three-dimensional state. Chaparones in these instances recognize an incorrectly unfolded protein and provide an energetically favorable pathway, through the hydrolysis of ATP, for the protein to unfold and refold to reach its functional state. Even in these cases, the structure of the protein is dictated by its amino acid sequence.

In principle, it should be possible to deduce the structure of a protein from its amino acid sequence. At this time, it is not possible to perform *ab initio* structure prediction with any great success. As such protein structure prediction remains one of the major problems in biology. Progress in structure prediction has been made through the combination of sequence and structural similarities. This offers hope that with the knowledge of sufficient structures across a wide range of organisms it should be possible to generate the structure of all unknown proteins. Although there is still much to be learned about protein structure, a series of fundamental features, folding rules, and structural motifs have been observed in many of the three-dimensional structures determined to date. These common features arise as a consequence of the amino acids used to build the proteins, the peptide bonds that join the amino acids, and the thermodynamic factors that control protein stability. These common threads in protein structure are described in the following.

II. AMINO ACIDS

All proteins are synthesized from the 20 α-amino acids specified by the genetic code as shown in Fig. 1. The nature of an amino acid is determined by the "side-chain" attached to the α-carbon (Table I). All of these amino acids, except for glycine which carries two hydrogens on its α-carbon, have a chiral center located at the α-carbon. Thus the amino acids exist as either the L- or D-isomers. Only the L-stereoisomer is utilized in protein biosynthesis (Fig. 2). This introduces chirality into all protein molecules that is the source of most of the asymmetric features found in protein structures. The use of only one of the two stereoisomers of the amino acids also establishes a structural uniqueness that is essential for biochemical specificity.

There are four classes of amino acids specified by the genetic code: (1) aliphatic amino acids, (2) aromatic amino acids, (3) polar amino acids, and (4) charged amino acids. These groups of amino acids provide the range of properties necessary to create a stable, functional folded protein.

As discussed elsewhere the primary driving force in protein folding and protein structure is the hydrophobic effect. This serves to sequester the hydrophobic side chains away from the bulk solvent. Once folded a typical protein is a densely packed entity that contains few holes larger than a water molecule. The aliphatic amino acids which include glycine, alanine, valine, leucine, isoleucine, and proline provide the range of small hydrophobic amino acids necessary to fill the gaps in the interior of the protein. Glycine and proline serve special roles in protein structure. Glycine is the smallest amino acid and is unique because it lacks a side chain. This gives it more conformational freedom than any other amino acid. Glycine is often found in turns and loops where other amino acids would be sterically unacceptable. It is also found where secondary structural elements intersect and other side chains would introduce molecular collisions. In contrast proline is unusual because it is conformationally restricted. As such it is often found in turns since it introduces an inherent kink in the polypeptide chain without any entropic cost to protein folding. Proline is also unique in that it is the only amino acid (or technically an "imino acid") that is commonly found to form a *cis* peptide bond between itself and the residue that precedes it in the polypeptide chain. In this instance the energy barrier to rotation is considerably less than all other peptide bonds (13 kcal/mol vs ~20 kcal/mol). This post-translational conformational modification often represents a slow step in protein folding.

Phenylalanine, tyrosine, and tryptophan are large aromatic residues that are normally found buried in the interior of a protein and are important for protein stability. Tyrosine has special properties since its hydroxyl side chain may function as a powerful nucleophile in an enzyme active site (when ionized) and is a common site for phosphorylation in cell signaling cascades. Tryptophan has the largest side chain and is the least common amino acid in proteins. It has spectral properties that make it the best inherent probe for following protein folding and conformational changes associated with biochemical processes.

The polar amino acids include serine, threonine, cysteine, methionine, asparagine, and glutamine. These are an important class of amino acids since they provide many of the functional groups found in proteins. Serine often serves as a nucleophile in many enzyme active sites, and is best known for its role in the serine proteases. Both serine and threonine are sites of phosphorylation and glycosylation which are important for enzyme regulation and

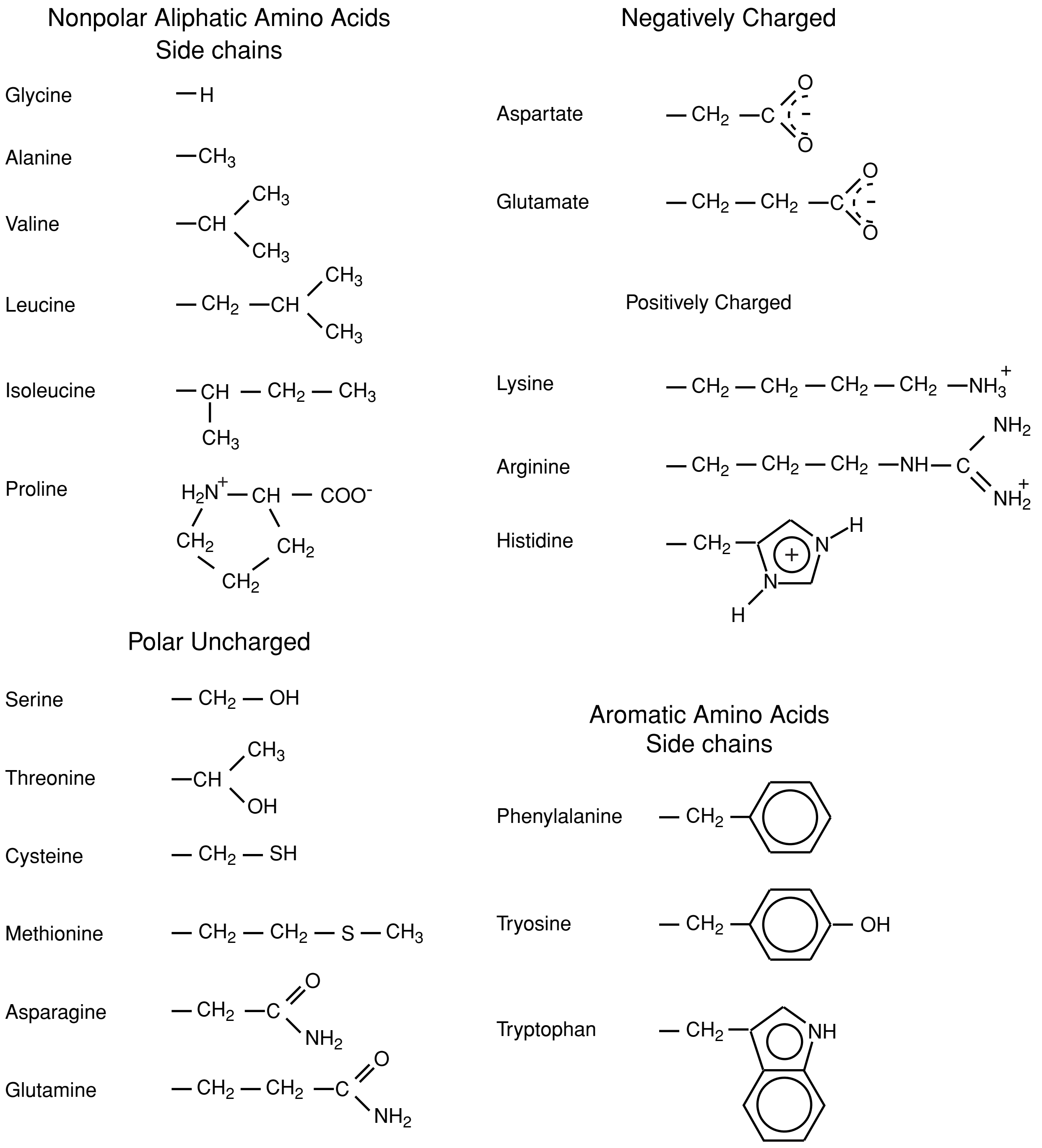


FIGURE 1 The 20 amino acid side chains specified by the genetic code. All except glycine have a β-carbon. Proline is technically an imino acid since it is a secondary amine.

cell signaling. Cysteine is the most reactive amino acid side chain. It serves as a potent nucleophile and metal ligand (particularly for iron and zinc), but is best known for its ability to form disulfide bonds, which often make an important contribution to the stability of extracellular proteins. Methionine is a fairly hydrophobic amino acid and typically found buried within the interior of a protein. It can form stacking interactions with the aromatic moieties of tryptophan, phenylalanine, and tyrosine. Asparagine and glutamine are close relatives of aspartate and glutamate but differ in the lack of charge and altered hydrogen bonding characteristics. In general these are not very reactive residues; however, asparagine is a common site for glycosylation.

TABLE I Properties of the Amino Acids

Amino acid			Side chain pK_a	Occurrence
Aliphatic				
Glycine	Gly	G		7.2
Alanine	Ala	A		7.8
Valine	Val	V		6.6
Leucine	Leu	L		9.1
Isoleucine	Ile	I		5.3
Proline	Pro	P		5.2
Aromatic				
Phenylalanine	Phe	F		3.9
Tyrosine	Tyr	Y	10.5	3.2
Tryptophan	Trp	W		1.4
Polar uncharged				
Serine	Ser	S	∼13	6.8
Threonine	Thr	T	∼13	5.9
Cysteine	Cys	C	8.4	1.9
Methionine	Met	M		2.2
Asparagine	Asn	N		4.3
Glutamine	Gln	Q		4.3
Positively charged				
Lysine	Lys	K	10.5	5.9
Arginine	Arg	R	12.5	5.1
Histidine	His	H	6.0	2.3
Negatively charged				
Aspartate	Asp	D	3.9	5.3
Glutamate	Glu	E	4.1	6.3

The charged amino acids include aspartate, glutamate, lysine, arginine and histidine. As a group these amino acids are relatively abundant and are important for making proteins soluble. Thus, these residues are generally located on the surface of the protein unless they play a specific biological role. Aspartate and glutamate are negatively charged amino acids. Both of these residues can function as general acids or bases in enzyme catalyzed reactions. Likewise they are important metal ion ligands. Lysine, arginine, and histidine can carry a positive charge. Of these, arginine is constitutively positively charged since its pK_a lies around 12.5. Lysine also plays an important role in coordinating negatively charged ligands; however, it functions as a nucleophile in some enzyme catalyzed reactions. Histidine is perhaps the most common and versatile catalytic residue in proteins. Its pK_a of ∼6.0 allows it to function both as a catalytic acid or base at physiological pH depending on its local environment. Histidine also has the ability to form covalent intermediates during catalysis such as phosphohistidine. In addition, it is often a ligand for transition metal ions such as iron and zinc.

A. Post-Translational Modifications

Once synthesized and folded, many proteins undergo post-translational modifications before they reach a functional state. Over 200 variant amino acid residues have been identified in proteins thus far. These changes are almost always achieved through an enzymatic pathway. The simplest changes include the formation of disulfide bonds (discussed later) and proteolytic processing of the polypeptide chain to yield a functional protein. Examples of proteolytic processing include the removal of signal peptides, the activation of zymogens to generate active forms of many proteolytic enzymes, and the maturation of viral proteins. Additionally proteolytic processing occurs in the biosynthetic pathway of many hormones. Other simple changes include the glycosylation of asparagine, serine, threonine, and phosphorylation of serine and tyrosine.

It is noteworthy that many post-translation modifications are associated with a sequence motif such that it is frequently possible to identify potential sites directly from the amino acid sequence. This arises because most post-translational modifications are the result of enzymatic pathways, which are usually highly specific. Thus protein sequences inferred from DNA sequence are often annotated with sites for post-translational modification. These sites should be viewed with caution since proof of

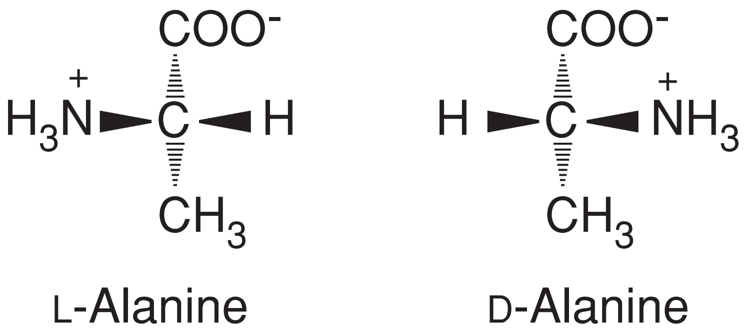


FIGURE 2 Stereoisomers of L-alanine and D-alanine.

modification can only be obtained through chemical or physical experimentation.

B. Cofactors

Although many proteins derive all of their function from their constitutive amino acids, a large number of proteins require additional cofactors in order to fulfil their biological role. These cofactors provide chemical properties that are not present in the 20 amino acid residues. For example, none of the amino acids are capable of facilitating an oxidation/reduction reaction. A wide range of cofactors are utilized including inorganic ions such as Fe^{2+}, Mg^{2+}, and Zn^{2+}, or complex organic molecules that are normally described as coenzymes such as flavin adenine dinucleotide or nicotinamide adenine dinucleotide. If the coenzyme is covalently bound to the enzyme it is often called a prosthetic group. The complete enzyme is called a holoenzyme, whereas the protein in the absence of its cofactors is called an apoenzyme or apoprotein. Many apoproteins are considerably less stable in the absence of their cofactor. This suggests that although the amino acid sequence dictates the overall three-dimensional structure, the cofactor is an integral part of the protein.

C. Context Determines Function

Typically proteins and especially enzymes contain only a few residues that are absolutely vital for function. In contrast there are usually many other residues of the same type in the protein that do not fulfill any special role. In many cases the catalytic residues have different chemical and physical properties from the same amino acid in solution; for example, the pK_a of the side chain might be several pH units higher or lower than the free amino acid. It is generally found that the behavior of an amino acid is profoundly influenced by the context of that amino acid within the protein. Altering the chemical properties of a functional group is one of the major attributes of protein structure and appears to be essential for the activity of most enzymes. There are many ways in which this is achieved; however, a simple example is placement of a charged residue in the interior of a protein such that the deionized state is favored. This serves to raise the pK_a of aspartate and glutamate and lower the pK_a of lysine.

III. PROTEIN STRUCTURE DETERMINATION

A major requirement for understanding protein structure is a large database of three-dimensional structures. This is particularly important for the comparative method of structure prediction. Although considerable progress has been made in recent years toward establishment of a comprehensive structural database many more protein models are needed before structures can be predicted with a high degree of confidence. There are two methods by which protein structures can be determined: X-ray crystallography and NMR. These techniques are complementary, with each having its advantages for providing information about specific aspects of protein structure. A detailed description of these methods is beyond the scope of this summary, but a few comments are noteworthy

A. X-Ray Crystallography

The first structure of a protein, myoglobin, was determined by X-ray crystallography in 1958 and was followed soon thereafter by the structure of hemoglobin. At that time protein structure determination was a daunting undertaking and few structures were determined in the ensuing years. Fortunately continual developments in the fundamental understanding of X-ray crystallographic theory, data collection, and computational methods have made the determination of protein structure routine. The result of this approach is an **electron** density map, which is interpreted in terms of a molecular model. The strength of this technique is that it can be applied to any macromolecular assembly that can be crystallized. The overwhelming majority of structures in the protein databank have been determined by X-ray crystallography.

The limiting factor in a successful X-ray structure determination is the growth of high quality crystals. In general if suitable crystals can be obtained a three-dimensional structure will be determined. The final quality of an X-ray structure is directly dependent on the three-dimensional order of the crystals since X-ray crystallography is an imaging technique. This is usually indicated by the "resolution" of the data. Resolution refers to the minimum diffraction spacing included in the structural determination where a smaller the number corresponds to a better structure. Typically a structure at 2.8 Å resolution is satisfactory to determine the path of the polypeptide chain, but data better than 2.5 Å are required to define the hydrogen bonding pattern in a protein with great confidence.

The one concern leveled at X-ray structures is the influence of the crystalline lattice on the observed conformation of the protein. Fortunately it has been demonstrated repeatedly that the structures of proteins observed in crystalline lattice are consistent with most of the biochemical measurements on the same protein. This arises because protein crystals typically contain about 50% solvent such that very little of a protein molecule is in contact with its neighbors in the crystal lattice and the packing forces are thermodynamically small. In some cases proteins are

enzymatically active in the lattice. In others conformational changes are observed between the substrate-free and substrate-bound forms of the enzyme. Typically this requires the crystallization of site-directed mutant proteins complexed with the substrate(s) or the study of complexes with substrate analogs. Except for the use of Laue techniques, protein crystallography yields a time-averaged view of the protein structure. Careful analysis of accurate X-ray diffraction data may provide some indication of conformational flexibility, but that aspect of protein structure is best suited to spectroscopic techniques such as NMR.

B. NMR

The use of NMR to determine protein structures is a more recent development than X-ray diffraction. It has the advantage that the analysis can be performed in the solution state of the protein which removes any artifacts introduced by crystallization. Its major disadvantage is the size limitation, which restricts most analyses to smaller proteins (<40 kDa), although it is anticipated that improvements in the technology will extend the size limitation.

Structural studies on proteins became possible with the advent of multidimensional NMR techniques. These rely on the use of isotopic labeling with ^{13}C, and ^{15}N and techniques to provide a facile method for assigning all of the ^{1}H resonances in a protein, which would otherwise be a difficult task. The measurement of nuclear Overhauser effect (NOE) intensities provide much of the distance information necessary to derive a structure, although additional chemical shift information is needed for a high-resolution structural determination.

Once a set of distance information has been obtained a series of models are generated and optimized by energy minimization and molecular dynamics within the restraints imposed by the distance information. The advantage of this approach is that it provides structural information on the protein in solution, the drawback is that surface residues and loops appear less well defined because there are generally fewer distance restraints for these components. The great strength of NMR is that it can yield specific information concerning the pK_a of an individual group in a protein as well as providing insight into the dynamical properties of the macromolecule.

IV. STRUCTURAL HIERACHY

The structure of a protein is generally understood in terms of an organizational hierarchy that consists of protein sequence, local secondary structure, tertiary structure, and finally quaternary structure. The study of protein structure in these terms has led to a greater understanding of the underlying physical principles that control the conformation and function of proteins. This hierarchy also reflects one conceptual view of protein folding, where the local secondary structural elements form first, followed by the tertiary and quaternary structure. The following discussion of protein structure is organized according to the preceding hierarchy.

A. Primary Structure

The character of a protein is determined by the amino acid sequence and composition of the polypeptide chain. By convention the order of amino acids in a protein is listed starting at the N-terminal and ending at the C-terminal amino acid residue. The N-terminal amino acid carries a free amino group, whereas the C-terminal residue retains a free carboxyl group. These terminal residues of the polypeptide chain are also referred to as the amino and carboxy terminus of the protein, respectively. Almost all protein sequences are determined indirectly by DNA sequencing. Chemical sequencing, either by automated Edman degradation or by mass spectroscopy, is still necessary to identify a protein from its original source and to prove the presence of post-translational modifications. All sequences of interest should be examined for errors by resequencing and comparison with orthologous proteins.

B. Conformational Restrictions on the Polypeptide Chain

The amino acids in a protein are linked by an amide linkage that is referred to as the peptide bond (Fig. 3). There are several key features to this bond. It is planar as a consequence of the partial double bond character of the carbon–nitrogen bond. It is almost always in the *trans* configuration. The peptide bond is fairly rigid where the barrier to rotation is ~20 kcal/mol. The carbonyl oxygen and amide hydrogen carry a partial negative and positive charge, respectively, which allow each of them to form a hydrogen bond. This linkage profoundly influences the stability, conformation, structure, and function of proteins.

The atoms that form the backbone of the linear polypeptide chain are usually referred to as the main-chain atoms. By convention the conformational (torsional) angles adopted by the main chain atoms are denoted by ϕ, ψ, and ω as shown in Fig. 3. Of these, ω describes the peptide bond and usually adopts a value of 180°. In principle there is free rotation about the other two angles (ϕ and ψ); however, the peptide bond and presence of a β-carbon places substantial restrictions on these conformational angles.

Conformational energy calculations and experimental observations based on high-resolution X-ray structure determination show that generally only ~8% of the possible

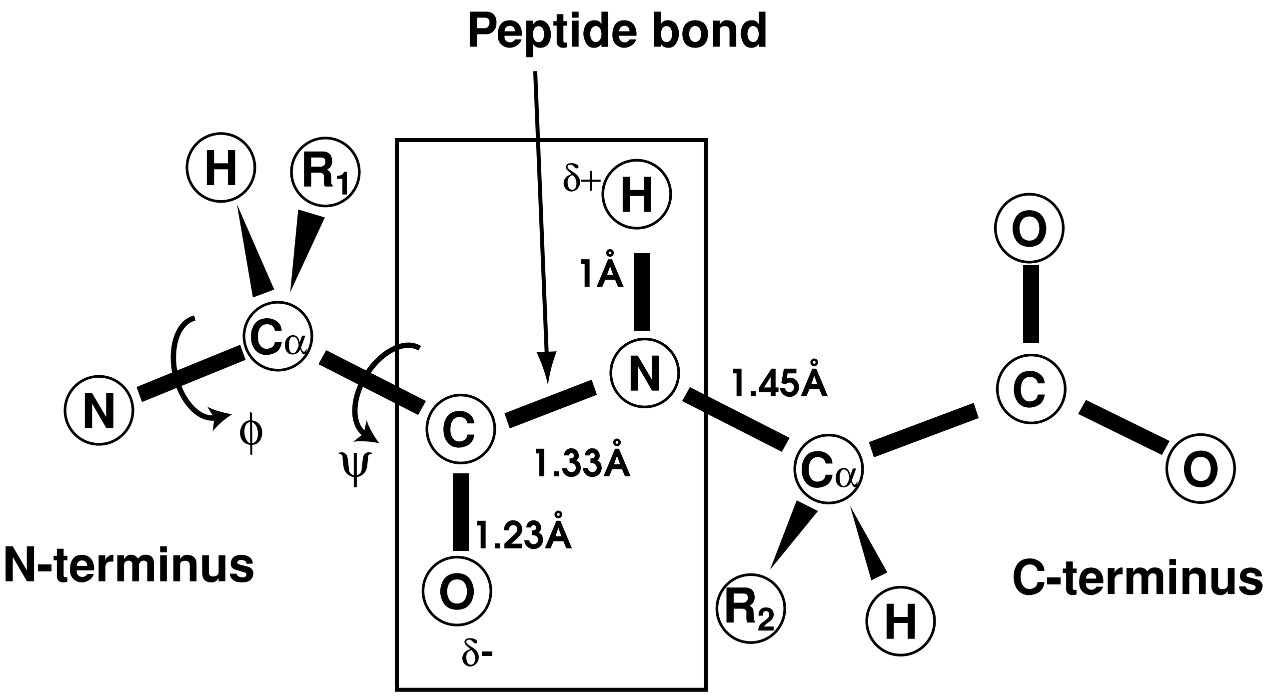


FIGURE 3 Schematic representation of the peptide bond and the observed restraints on the conformational.

combinations of ϕ and ψ are strictly allowed, whereas more generous energy considerations include a total of 20% (Fig. 4). These allowed regions of conformational space fall into three areas that are occupied by the major secondary structural motifs observed in protein structures. As indicated these belong to right- and left-handed α-helices and β-sheet.

V. SECONDARY STRUCTURAL MOTIFS

The major three-dimensional motifs found in proteins were predicted to exist by Cory and Pauling in 1951 before the first protein structure determination through their study of the structures of small peptides. They recognized that secondary structural motifs must accommodate the hydrogen bonding potential of the peptide bond as well as utilize the conformational angles found in model peptides. This emphasizes the importance of hydrogen bonds in specifying the conformation of the polypeptide chain. In general every potential hydrogen bond donor and acceptor in a protein participates in one or more hydrogen bonds. This requirement explains the common occurrence of the α-helix and β-sheet.

A. α-Helix

The first secondary structural element predicted and identified in a protein was the α-helix (Fig. 5). In the α-helix the path of the polypeptide chain follows a right-handed arrangement where carbonyl oxygen of residue (i) interacts with the amide hydrogen on residue ($i + 4$). There are 3.6 residues per turn with a helical rise of 1.5 Å per residue which gives a helical pitch of 5.4 Å. As a consequence the side chains extend away from the helix axis every 100°. The side chains also extend toward the N-terminus of the α-helix due to the chirality of the amino acids. This is a very compact arrangement of residues that satisfies the hydrogen bonding requirements of the polypeptide chain except for the four amide hydrogens at the N-terminus and four carbonyl oxygens at the C-terminus of the helix. The lack of hydrogen bonding at the ends of an α-helix explains why they always contain more than one turn.

The length of α-helices varies enormously from a few turns in globular proteins to hundreds as seen in extended proteins such as myosin. Keratin is one of the most abundant fibrous proteins and is almost entirely α-helical in nature. In general there are some preferences for those amino acid residues found in α-helices that prove useful for qualitative structure prediction. For example, proline is commonly found at the N-terminus of a helix, but rarely found in the middle. Not only does the proline lack an amide hydrogen but also the pyrolidine ring restricts the preceding residue from adopting the conformational angles necessary for helix formation.

As indicated in the Ramachandran plot (Fig. 4), the left-handed α-helix is an allowed conformation. It is observed

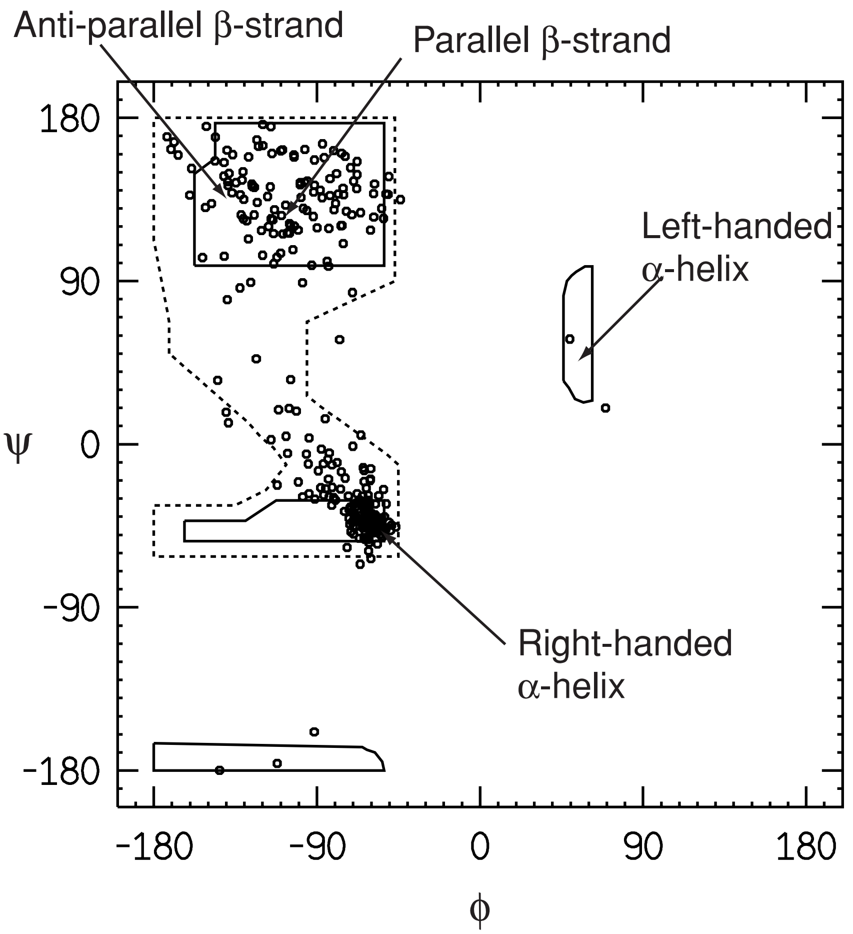


FIGURE 4 Ramachandran plot for a high-resolution protein structure (human UDP-galactose 4-epimerase). This reveals that only a limited part of conformational space is occupied by the main-chain torsional angles. The space enclosed by the solid lines represents the most energetically favorable regions of conformational space. The dashed lines indicate more generously allowed regions.

occasionally in proteins, but not for an extended number of residues on account of unfavorable interactions between side chains. Residues that adopt this conformation are usually located in turns. The dominance of the right-handed helical conformation over the left-handed is a direct consequence of the L-amino acids.

One other helical conformation is found in globular proteins: the 3_{10} helix. This differs from the α-helix by the location of the hydrogen bond. In the 3_{10} helix a hydrogen bond is formed between the $C{=}O_{(i)}$ and $H{-}N_{(i+3)}$. The packing of the main-chain atoms in the 3_{10} helical conformation is quite tight thereby yielding nonlinear hydrogen bonds and thus is not found for extended periods. The only location that 3_{10} helical conformation is fairly common is at the C-terminal ends of α-helices where it serves to terminate the helix with a tight turn. Three residues in the 3_{10} helical conformation constitute a type III turn. In principle the π-helix which has a hydrogen bond between $C{=}O_{(i)}$ and $H{-}N_{(i+5)}$ is an allowed conformation. It is not observed in proteins because the packing of the main-chain atoms would be too loose giving rise to a hole through the center of the helix. In addition there is steric hindrance between the side chains of adjacent residues along a π-helix.

B. Collagen Helix

Collagen is the most abundant protein in mammals. It is a fibrous protein that exhibits a helical repeat that is different from that of any of the previous conformations. This protein is characterized by a repeating motif $(Gly{-}X{-}Y)_n$ where X is usually proline and Y is often 4-hydroxyproline. Each collagen molecule contains ~1000 amino acid residues and is about 3000 Å long. This protein is synthesized as a preprotein that includes 200 additional residues at both the N- and C-terminii that fold to form globular domains and serve to prevent the molecules

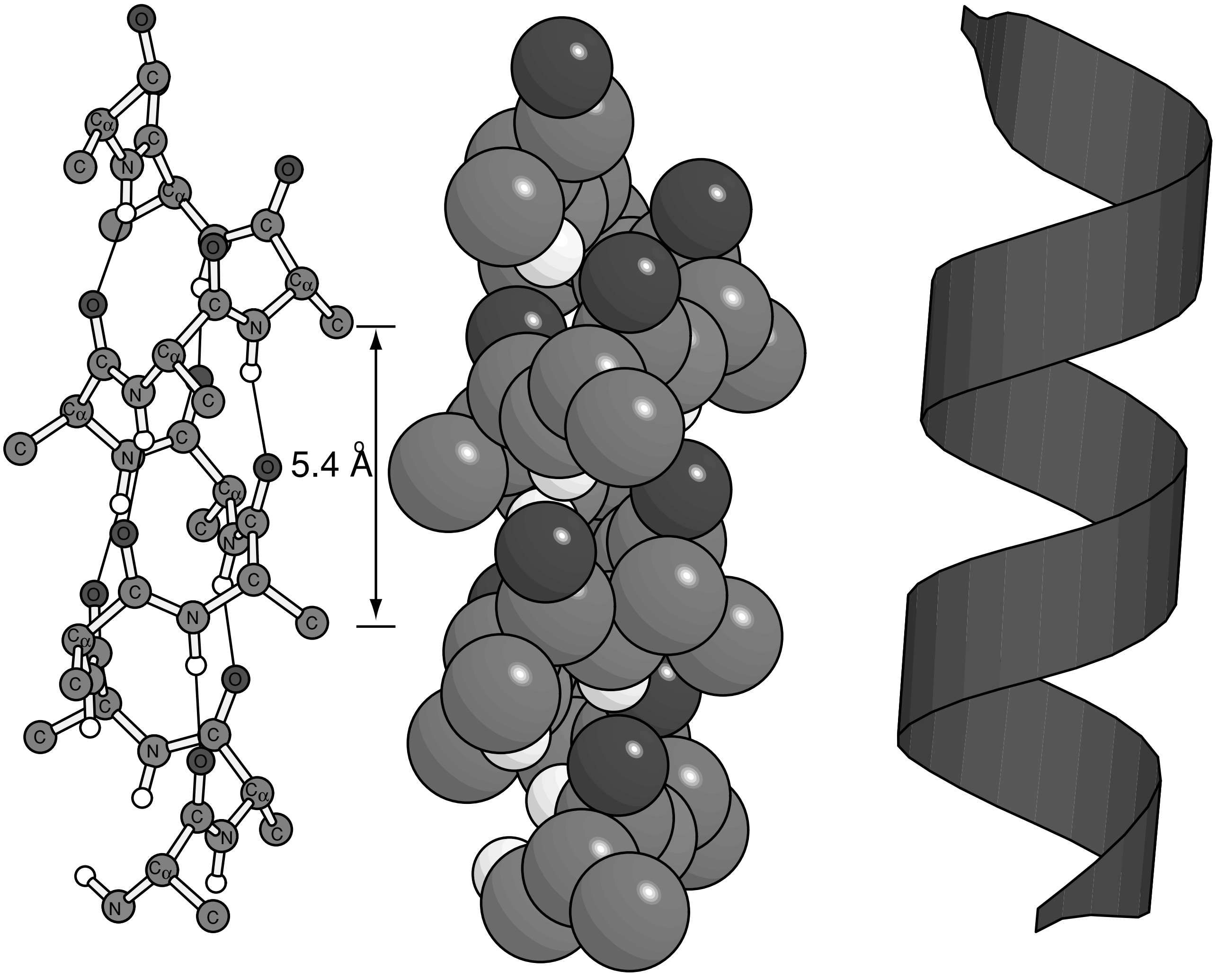


FIGURE 5 The α-helix (a) ball and stick representation, (b) space filling representation, and (c) ribbon representation.

from assembling. After post-translational hydroxylation of approximately one half of the prolines and removal of the globular extensions at both termini each collagen protomer adopts a left-handed polyproline (II) helical conformation and assembles with two other molecules to form a right-handed triple-helix (Fig. 6). Each strand of collagen is highly extended with 3.3 residues per turn and a translation of 2.9 Å per residue. Consequently individual strands are unstable and must aggregate for stability. The noninteger number of residues per turn is to accommodate the right-handed superhelix that contains 10 triplets per turn. With this arrangement, every third residue lies at the helix axis in such a way that only a glycine would fit (Fig. 6). This explains the $(Gly{-}X{-}Y)_n$ sequence motif. Interestingly there is extensive hydrogen bonding in the collagen superhelix; however, it occurs between the carbonyl oxygen on one strand with a NH on a neighboring molecule rather than within the polypeptide as seen in the α-helix. The use of proline and hydroxyproline restricts the conformational angles available to the polymer and serves to stabilize the collagen fibers. Hydroxylation of the proline increases the stability of the fibers, although the exact reason for this enhanced stability is still unclear.

C. β-Sheet

The second most common and identifiable secondary structural conformation is the β-strand. In contrast to the α-helix, the polypeptide chain in a β-strand is almost completely extended with a translation per residue of 3.4 Å. An isolated strand is unstable because there are no interactions between residues that are close in sequence. Thus the

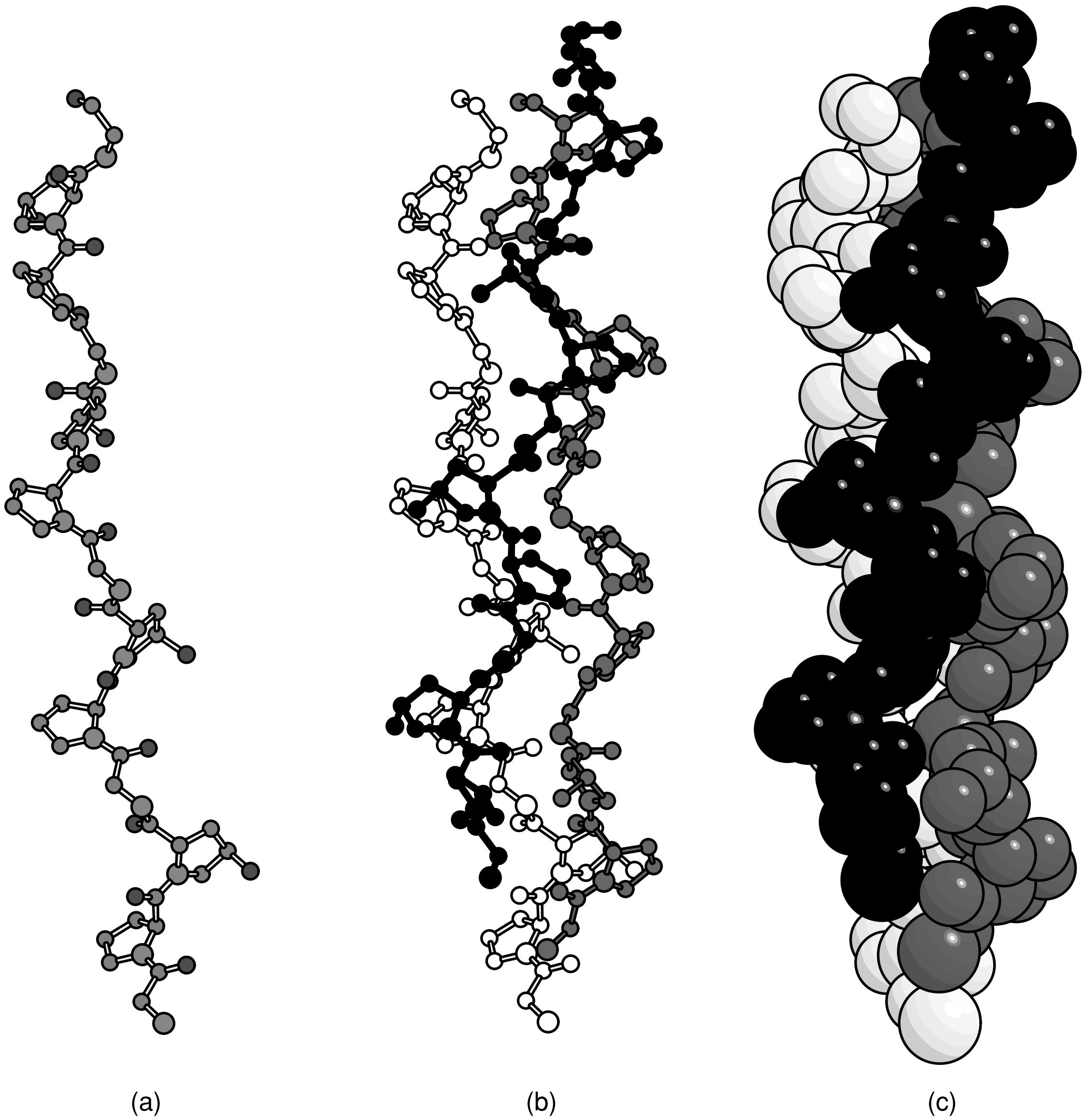


FIGURE 6 The collagen helix (a) single strand of Gly–Pro–Hyp, (b) triple helix of collagen, (c) space filling representation of the collagen triple helix.

β-strand is only observed in conjunction with other strands where it can form complementary hydrogen bonds with opposing peptide groups. These strands can either associate in a parallel or antiparallel manner to form β-sheets. Association of multiple strands gives rise to larger sheets that may be built from a variety of antiparallel and parallel strands; however, there is a strong tendency to prefer structural motifs that are dominated by mostly parallel or antiparallel strands.

In both antiparallel and parallel strands the peptidyl oxygen and amide hydrogen form almost ideal hydrogen bonds with neighboring strands; however, the geometry is somewhat different in each type of sheet (Fig. 7). Both arrangements lead to stable structures; however, antiparallel β-sheets are generally considered to be more stable than sheets built solely from parallel strands. As discussed later parallel sheets typically are buttressed on both sides by additional layers of secondary structure where these are usually α-helices. In contrast anti-parallel β-sheets often only require one additional layer of secondary structural elements to establish a stable fold.

The side chains in both parallel and antiparallel β-sheets extend alternatively to opposite sides of the sheet. Consequently the groups on adjacent residues within a strand do not contact each other. Rather there is considerable interaction between side chains on adjacent strands

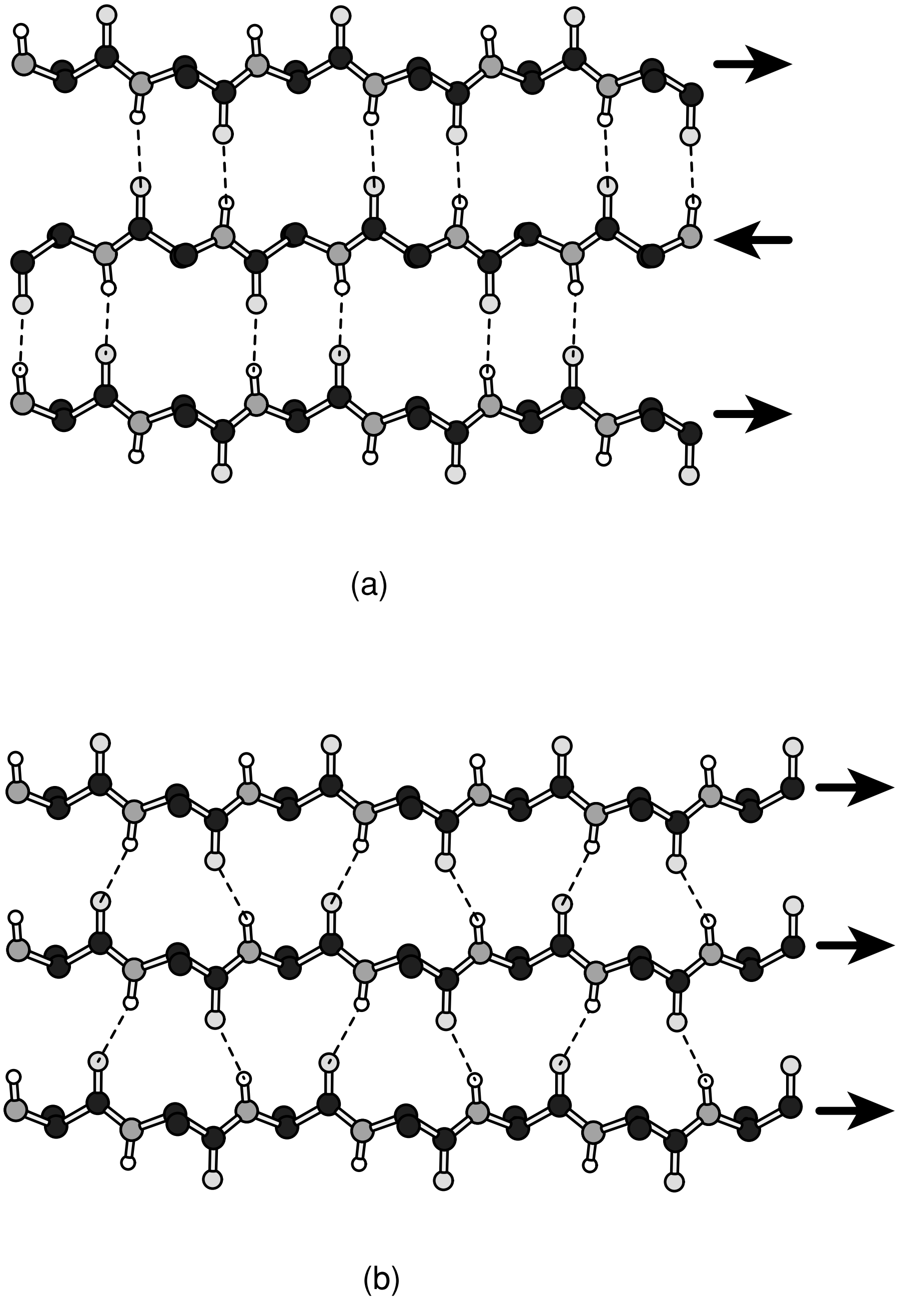


FIGURE 7 Ball-and-stick representations of (a) antiparallel and (b) parallel β-sheets. By convention the direction of the polypeptide chain is taken to run from the N-terminus to the C-terminus.

such that complementarity is observed. In addition, all β-sheets have a characteristic right-handed twist when viewed along the strand. This twist is considered to be the consequence of the interactions of the side chains with the backbone of the polypeptide chain and is thus a direct result of the chirality of the amino acids. The magnitude of the twist is somewhat variable but is usually more pronounced in antiparallel β-sheets.

There is a considerable contrast between the nature and usage of the α-helix and β-sheet. α-Helices are self-contained secondary structural elements that may contain a substantial number of amino acid residues even in globular proteins. By comparison, β-strands typically contain three to six amino acid residues and require an adjacent strand to form a stable folding unit. Proteins that are built from α-helices usually have a very high percentage of their

residues in the helical conformation (~80%) with comparatively few devoted to connecting regions. In proteins that are dominated by β-strands, typically <50% are in the β-conformation. This occurs because for every three to six residues in each strand there must be an equivalent number of amino acids devoted to a turn to bring the polypeptide chain back into a position where it can hydrogen bond to the same or neighboring β-strand. This emphasizes the importance of turns in protein structures.

D. Turns and Random Coil

Many proteins contain secondary structure that cannot be described as either helix or turn. This is typically classified as turn, loop, or random coil. These sections of the polypeptide chain are characterized by nonrepetitive conformational angles; however, this does not necessarily imply that these residues are less stable or less well ordered than the regular secondary structural elements. Many active site residues and components critical for ligand recognition reside in loops or random coil and adopt an exquisitely well-defined conformation.

On average, one third of all residues in proteins are involved in turns that serve to reverse the direction of the polypeptide chain. These turns are an essential feature of globular proteins and are almost always located at the surface. In contrast to α-helices and β-strands which have repetitive conformational angles, the conformational angles observed in turns occur in sets that are characteristic of each type (Table II). Turns have been classified according to the commonly observed groups of conformational angles and the number of residues involved. Of these the β-hairpin or reverse turn is the most common. This type of turn is frequently used to connect antiparallel β-strands.

Three general types of reverse turn have been described; types I, II, and III, which all contain four amino acid residues and normally exhibit a hydrogen bond between $C{=}O_{(i)}$ and $H{-}N_{(i+3)}$ (Fig. 8). Of these, the type III turn consists of a short section of residues in the 3_{10} helical conformation. Additional variants of the type I and II class are observed in the I′ and II′ turns. These exhibit conformational angles for the central two residues of the turn that are the mirror image of types I and II. The observed conformation angles favor the presence of certain amino acids at specific locations in the turns. For example, glycine predominates at position $(i + 3)$ and proline predominates at position $(i + 1)$ in both types I and II turns. In all turns the central two amino acid residues do not form peptidyl hydrogen bonds within the turn itself and thus must either accommodate their hydrogen bonding potential via a side chain interaction with a neighboring residue or through interactions with the solvent. Thus polar or charged residues (Asp, Asn, Ser) are often located at the first residue of the turn so that they can form a hydrogen bond to the amide hydrogen of residue $(i + 2)$. The need to satisfy the hydrogen bonding potential of the main chain atoms accounts for the placement of most turns at the surface of the protein.

VI. PROTEIN STABILITY

The term protein stability refers to the energy difference between the folded and unfolded state of the protein in solution. Remarkably, the free energy difference between these states is usually between 20 and 80 kJ/mol, which is of the magnitude of one to four hydrogen bonds. Although this suggests that proteins are only marginally stable, the stability is sufficient to prevent spontaneous unfolding at normal temperatures.

TABLE II Conformational Angles of the Major Secondary Structural Elements

Secondary structure		ϕ	ψ		
Helical conformations					
	α-Helix	−57	−47		
	3_{10} Helix	−49	−26		
	Collagen Helix	−78	+149		
β-Strands					
	Antiparallel	−139	+135		
	Parallel	−119	+113		
β-Turns		$\phi(i+1)$	$\psi(i+1)$	$\phi(i+2)$	$\psi(i+2)$
	Type I	−60	−30	−90	0
	Type I′	+60	+30	+90	0
	Type II	−60	+120	+80	0
	Type II′	−60	−120	−80	0
	Type III	−60	−30	−60	−30

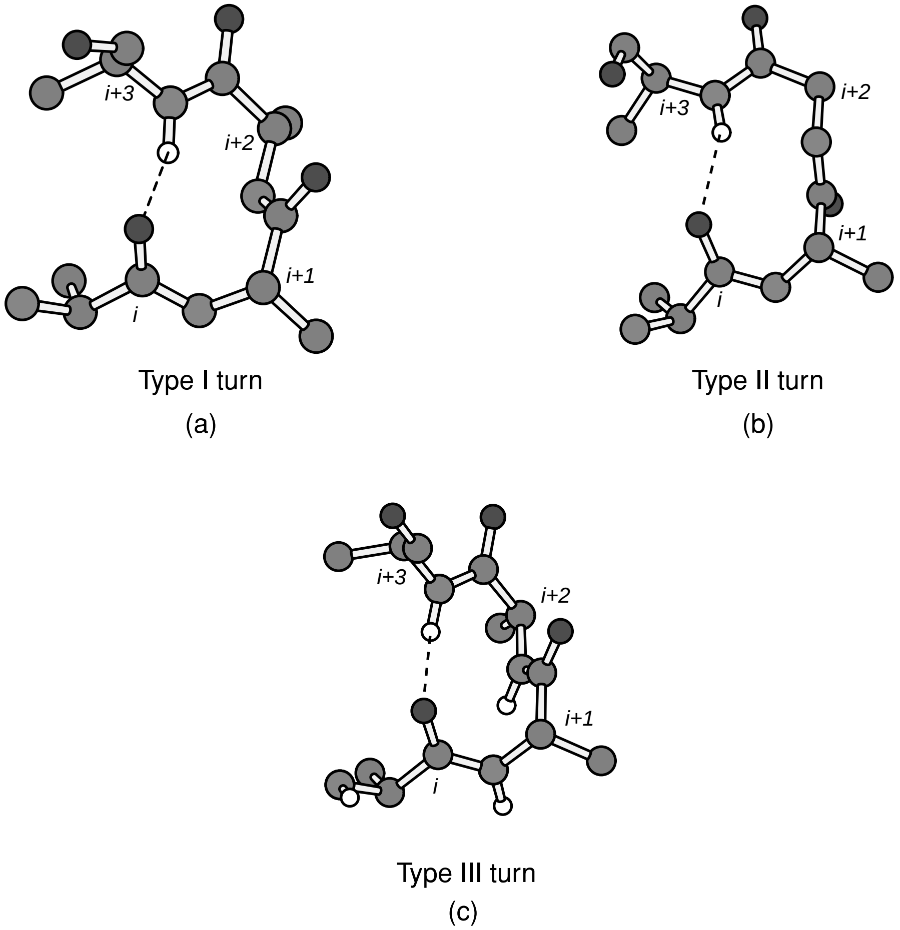


FIGURE 8 Types I, II, and III turns.

Protein stability is determined by an enormous number of weak interactions observed in the folded state which have to be balanced against an almost equivalent set of interactions with water in the unfolded state. This is a complex problem since every amino acid residue has potential for polar interactions via the peptide bond and a variety of ionic, polar, and nonpolar interactions through its side chains. This accounts for the difficulty in predicting structure directly from its amino acid sequence since the errors in any energy computation are far larger than the net stability of the protein.

A. Hydrophobic Effect

The major driving force in protein folding is the hydrophobic effect. This is the tendency for hydrophobic molecules to isolate themselves from contact with water. As a consequence during protein folding the hydrophobic side chains become buried in the interior of the protein. The exact physical explanation of the behavior of hydrophobic molecules in water is complex and can best be described in terms of their thermodynamic properties. Much of what is known about the hydrophobic effect has been derived from studying the transfer of hydrocarbons from the liquid phase into water; indeed the thermodynamics of protein folding closely follow the behavior of simple hydrophobic molecules in water.

Studies of hydrocarbon models demonstrate that at room temperature the insolubility of hydrophobic compounds is dominated by entropic rather than enthalpic considerations. This is often explained as a water-ordering effect where the insertion of a hydrophobic molecule into an aqueous environment induces a diffusely ordered water shell surrounding the molecule, akin to the formation of clathrates around noble gases and simple hydrocarbons. This shell forms because the hydrophobic compound cannot form hydrogen bonds to the water that surrounds it. Consequently those water molecules have a more restricted set of neighbors with whom to fulfil their hydrogen bonding capacities. This reduces their degrees of rotational freedom and thus leads to a reduction in entropy. This simple explanation based on ordering of water is

inadequate to fully explain the behavior of hydrophobic molecules in solution since both the standard changes in the entropy and enthalpy for the transfer of a hydrophobic molecule to water are strongly temperature dependent. Surprisingly the value for the free energy change for the transfer of a hydrocarbon to water is rather temperature insensitive as a consequence of compensatory changes in the entropy and enthalpy. These temperature dependencies are a consequence of the large temperature-insensitive heat capacity of hydrophobic molecules in solution. The water-ordering effect appears to be the source of the anomalously high heat capacity of hydrophobic compounds in water.

Although there is no general agreement on the molecular basis of the hydrophobic effect, there is a good correlation between free energy of transfer of an organic molecule to water and its hydrophobic surface area, whereas there are no general correlations between the molecular features of the solute and entropic and enthalpic changes. Many studies have shown that the transfer of 1 $Å^2$ of hydrophobic area to water is accompanied by an unfavorable increase in the free energy of 80–100 J/mol. The fact that this correlation is found to be fairly independent of the nature of the hydrophobic solute clearly indicates that the hydrophobic effect is a fundamental property of water.

Studies on the heat capacities changes observed at the protein folding transition show that protein denaturation is analogous to the transfer of hydrophobic molecules to water. Furthermore it is well established that the stability of a protein is directly proportional to the difference between the exposed hydrophobic surface area in the unfolded and folded state. In recent years, site-directed mutagenesis has demonstrated the same thermodynamic relationship between the hydrophobic buried surface area and stability as observed for the transfer of organic hydrocarbons into water. That is, each buried 1 $Å^2$ of hydrophobic surface area contributes ~80 J/mol to the stability of the protein when the only difference is the change in surface area. Clearly the behavior of a protein in solution is more complex than that of a simple hydrocarbon. As noted earlier, proteins are only marginally stable. It would appear that the change in exposed hydrophobic surface area that accompanies protein folding slightly more than compensates for the decrease in entropy of the polypeptide chain as it adopts a well-defined conformation. This explains why it is so difficult to predict protein structure.

B. Hydrogen Bonds

Hydrogen bonds (D—H····A) are primarily electrostatic in nature and involve an interaction between a hydrogen attached to an electronegative atom (D—H) and another electronegative acceptor atom (A) that carries a lone pair of electrons. In biological systems the electronegative atoms in both cases are usually nitrogen or oxygen. The distance between the donor and acceptor atoms is usually in the range 2.8–3.1 Å where the D—H bond tends to be collinear with the lone pair of electrons. There is some variability in the geometry of the hydrogen bond, which is consistent with the predominantly electrostatic nature of this interaction. For example, in the α-helix and antiparallel β-sheet the N—H is approximately colinear with the C═O bond rather than be aligned with the lone pairs of the carbonyl oxygen. Many of the hydrogen bonds in proteins occur in networks where each donor participates in multiple interactions with acceptors and each acceptor interacts with multiple donors. This is consistent with the ionic nature of hydrogen bonds in proteins.

Originally it was believed that hydrogen bonds made an important contribution to protein stability on account of the extensive hydrogen bonding observed in α-helices and β-sheets. Indeed, virtually every hydrogen donor and acceptor in a protein are observed to form an interaction within the folded structure or to the external solvent. However, protein stability is the difference in free energy between the unfolded state and the folded state. In the unfolded state the polar components are able to form perfectly satisfactory hydrogen bonds to water that are equivalent to those found in the tertiary structure of the protein. Thus hydrogen bonding is energetically neutral with respect to protein stability, with the caveat that any absences of hydrogen bonding in a folded protein are thermodynamically highly unfavorable.

Although hydrogen bonds do not contribute to stability they are a major determinant of protein conformation. The necessity to form hydrogen bonds accounts for the α-helices and β-strands that abound in protein structures.

C. Disulfide Bonds

Many extracellular proteins contain disulfide bonds. In these proteins the presence of disulfide bonds adds considerable stability to the folded state where in many cases reduction of the cystine linkages is sufficient to induce unfolding. The source of the stability appears to be entropic rather than enthalpic. The introduction of a disulfide bond reduces the entropy of the unfolded state by reducing the degrees of freedom available to the disordered polypeptide chain. This stabilizes the folded state by decreasing the entropy difference between the folded and unfolded state. This suggests an obvious strategy for increasing the stability of a protein through the introduction of disulfide bonds. Although this might seem a simple task, the geometry of the disulfide bond is rather restricted. As a consequence the number of locations that can accommodate the replacement of two residues by cysteines in a protein

without reducing the thermal stability of the folded protein are quite limited. Studies on T4 lyszoyme have shown that if suitable locations can be found, the degree of stability introduced into a protein is proportional to the size of the closed loop generated by forming a disulfide bond.

D. Ionic Interactions

The association of two oppositely charged ionic groups in a protein is known as a salt bridge or ion pair and is a common feature of most proteins. Typically these interactions contribute very little to protein stability since the isolated ionic groups are so effectively solvated by water. As a consequence very few unsolvated salt bridges are found in the interior of proteins. Furthermore, salt bridges are rarely conserved in orthologous proteins.

E. Dipole–Dipole Interactions

Dipole–dipole interactions are weak interactions that arise from the close association of permanent or induced dipoles. Collectively these forces are known as Van der Waals interactions. Proteins contain a large number of these interactions, which vary considerably in strength.

The strongest interactions are observed between permanent dipoles and are an important feature of the peptide bond. In the peptide bond the dipoles associated with the peptide carbonyl and amide group are aligned and give rise to a significant dipole moment (3.5 Debye units for a peptide bond *versus* 1.85 for a water molecule). These interactions fall off with the inverse of the second to third power when the dipoles are fixed and to the sixth power when they are free to rotate. So, for example, there is a substantial positive dipole at the amino-terminal end of an α-helix where the dipoles are constrained and aligned. As a consequence the N-terminal end of an α-helix is often utilized to bind negatively charged ligands in enzyme active sites.

Permanent dipoles may also induce a dipole moment in a neighboring atom or group. This is a stabilizing interaction, but is much weaker than that observed between permanent dipoles. These type of interactions are important since they change the charge distribution of neighboring atoms which in turn can profoundly influence activation barriers in enzyme catalyzed reactions.

London or dispersion forces are the weakest of all of the dipole–dipole. These are best described in quantum mechanical terms, but may be viewed qualitatively as the consequence of the transient asymmetry in the charge distribution in a neutral atom that induces a favorable dipole in a neighboring neutral atom thus leading to a weak attraction. These forces are inversely proportional to the sixth power of the separation. Although these forces are very weak there are an enormous number present within a folded protein such that they to contribute significantly stability of the folded state. As a group, the Van der Waals forces are important for stabilizing interactions between proteins and their complementary ligands whether the ligands are proteins or small molecules.

VII. TERTIARY STRUCTURE

A. Protein Folding Rules

Examination of a large number of protein structures has yielded a few common rules about the folds that proteins can adopt as listed in the following. The theoretical basis for these features is not well understood, but most appear to result from the chirality of the amino acids, entropic considerations, and the necessity to establish a hydrophobic core.

1. Secondary structural elements that are close in the sequence of a protein are often adjacent in the folded protein. It is less common to find secondary structural elements that are far apart in the sequence and close together in the structure. The exception to this arises where an auxiliary domain has clearly been inserted into a loop in a protein. This is probably the most entropically favorable way to arrange secondary structural elements within a folded protein.
2. Adjacent parallel β-strands are almost exclusively connected by right-handed crossovers (Fig. 9). It is believed that this feature arises from the chirality of the amino acids that leads to a net right-handed twist in the polypeptide chain.
3. There are no topological knots in proteins.
4. Proteins always contain more than one layer of secondary structural elements. This rule arises because proteins always contain a hydrophobic core formed by the association of hydrophobic side chains.
5. α-Helices and β-sheets typically associate in discrete layers of the same type of secondary structural elements. This feature is the consequence of the necessity to fulfill the hydrogen bonding requirements of the polypeptide chain and packing considerations. Typically the interior of a protein does not contain any holes larger than a water molecule. Because α-helices and β-strands differ greatly in their cross-sectional diameters, inclusion of these in the same layer would result in a poorly packed protein interior. In addition a mixture of α-helices and β-strands in the same layer would not fulfill the hydrogen bonding potential of the β-strand.

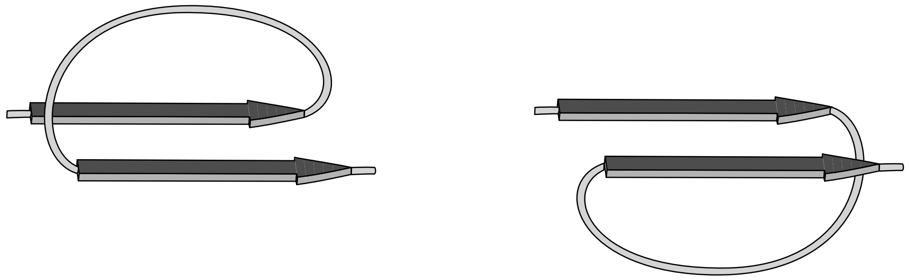


FIGURE 9 A right-handed cross-over connection joining two parallel β-strands. Note the right-handed twist of the β-strands when viewed along the strand axis.

B. Folding Motifs

Many protein structures are dominated by a few simple folding motifs. These represent thermodynamically favorable arrangements of secondary structural elements. These include the $\beta\beta$, $\beta\alpha\beta$, and $\alpha\alpha$ motifs as illustrated in Fig. 10.

The $\beta\beta$ and $\beta\alpha\beta$ motifs are commonly used to connect antiparallel and parallel β-strands, respectively. The $\beta\beta$ motif is frequently connected by a hairpin turn, which provides a compact way of changing the direction of the polypeptide chain. In the same way, the $\beta\alpha\beta$ motif provides a compact module where the width of the α-helix is similar to that of the combined width of the two β-strands. It also provides a hydrophobic core. The dimensions of the $\beta\alpha\beta$ motif explain why large parallel sheets that are built with this motif always have α-helices on both sides since there is insufficient space on one side of a sheet to accommodate all of the connecting helices.

A variety of $\alpha\alpha$ motifs are found in proteins depending on whether the α-helices are in contact with each other after the connecting loop. In cases where the α-helices are in contact they are typically inclined at an angle of either 20 or 50° reflecting the optimal ways to interdigitate side chains at their intersection. Both types of interaction are abundant in proteins and give rise to parallel or crossed helical bundles. There are also many important examples of $\alpha\alpha$ motifs where the connections between the two helices are longer to create a ligand binding site. Important examples of this type of motif are the helix–turn–helix motifs found in calcium binding proteins and DNA binding proteins.

C. Domains

The tertiary structure of a protein describes the manner in which the secondary structural elements are arranged in three dimensions to create a stable molecular entity. In many cases it is convenient to describe a protein in terms of regions of the polypeptide chain that might fold autonomously. These regions are called domains and much of the discussion of tertiary structure centers on classification of these units of protein structure.

Domains in proteins take on many forms. On some occasions it is clear that domains are connected by flexible hinge regions and that the domains could be expressed independently. In other cases the domains are built from apparently distant segments of the protein sequence such that it would be difficult to express those domains without rearrangement of the DNA. This illustrates an important difference in the use of "domain" in structural and molecular biology, since in the latter the term usually indicates a linear section of DNA that appears to influence a biological property where as in structural biology it represents an three-dimensional entity.

D. Protein Folds

Structural studies on proteins have uncovered a very wide variety of protein folds. At this time the upper limit of the number of unique ways in which proteins can fold is unknown; however, genomic sequencing has provided a limit for the maximum number of folds that might be needed for the life of an organism by providing an upper limit to the number of proteins in the genome. Fortunately, the number of unique folds is likely to be considerably less than the total number of proteins since many proteins of dissimilar function have been found to contain the same fold.

The assortment of protein folds observed thus far, at first glance, appears bewilderingly complex. Careful analysis of the common structural and topological features of these structures has lead to a classification of protein folds according to the content and arrangement of

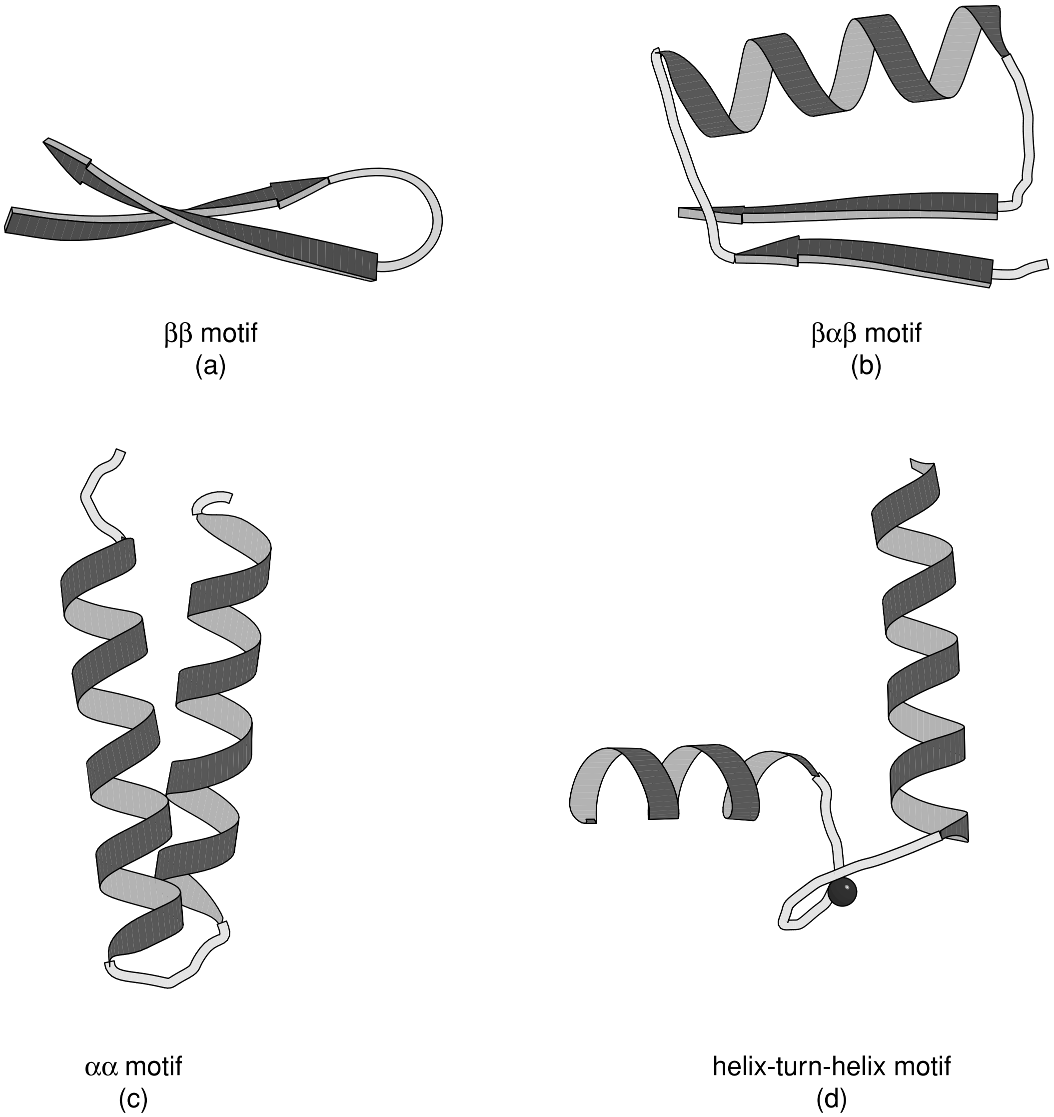
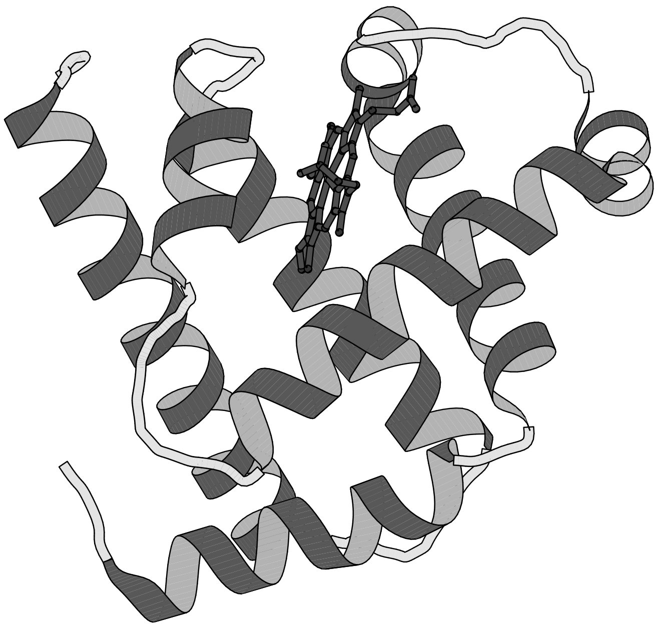


FIGURE 10 Ribbon representations of the (a) $\beta\beta$, (b) $\beta\alpha\beta$, (c) $\alpha\alpha$ motif, and (d) helix-turn-helix motif. Various forms of the $\alpha\alpha$ motifs are found depending on the manner in which the α-helices associate. (c) Shows the alignment of two helices joined by a short connection. (d) Shows the helix–turn–helix motif associated with calcium binding proteins.

the α-helices and β-strands. In turn, this has provided insight into the common underlying principles of protein structure. Several important databases exist of protein structures and tertiary structure classification. These include the RSCB (//www.rcsb.org/pdb/), CATH classification (//www.biochem.ucl.ac.uk/bsm/cath/), and SCOP (//scop.mrc-lmb.cam.ac.uk/scop/) structural databases. The first includes all of the coordinates for structures that have been made publicly available. The second two databases contain structural classifications for all the protein deposited in the RSCB. Both of these systems initially classify proteins into five major groups: all α, all β, α/β (where these secondary structural elements alternate through the fold), $\alpha + \beta$ (where the sections containing these secondary structural elements are segregated), and small proteins that are stabilized by metal ligands or disulfide bonds. Additional classifications have been added to incorporate multidomain proteins, membrane proteins, and peptides. Representative members of these families are discussed in the following.

E. All α-Proteins

Proteins that fall into this class typically consist of predominately α-helices (>60%), but may contain a small amount of β-sheet at their periphery (See CATH Classification). Historically the first two protein structures determined, myoglobin and hemoglobin, belonged to the all-α class.



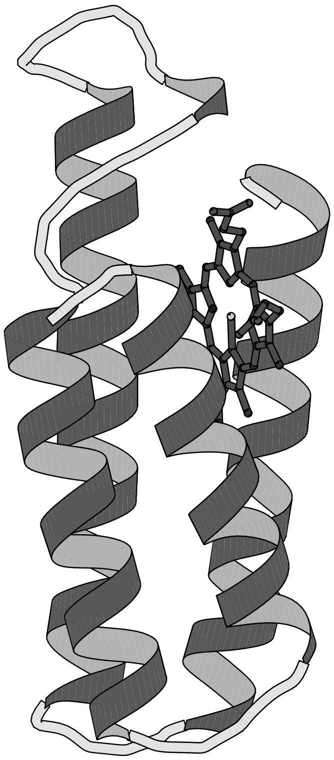


FIGURE 11 Ribbon representations of (a) myoglobin and (b) cytochrome B562. These are representative examples of all α-proteins which exhibit the two major ways of packing α-helices in proteins.

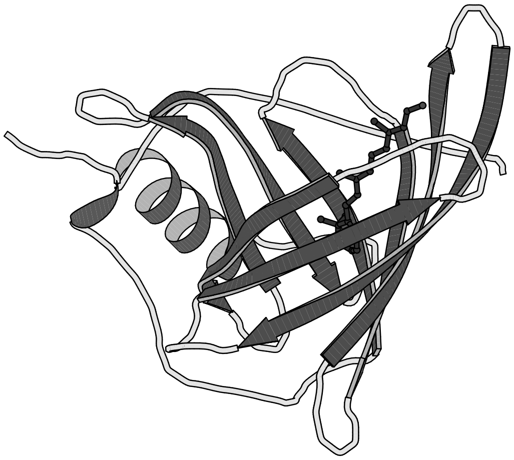
These are representative members of the crossed α-helical bundle motif and demonstrate one of the effective ways of packing helices into a protein core (Fig. 11a). The four-helix bundle illustrated in Fig. 11b shows the second way in which helices associate. When the connecting loops are short the packing leads to an antiparallel arrangement of helices; however, mixtures of parallel and antiparallel helices are also observed in folds that contain longer intervening sequences.

F. All β-Proteins

There are a large number of diverse protein folds that fall into the all-β class, but they all contain ~50% β-strand with only a very small amount of α-helix. Several representative examples of these are shown in Fig. 12. In structures composed primarily of β-strands there is always more than one layer which is necessary to establish a hydrophobic core. This means that at the edge of the sheets there must be some form of compensation for the disruption of the hydrogen bonding pattern. Most of the folds in this class are formed from antiparallel arrangements of β-strands.

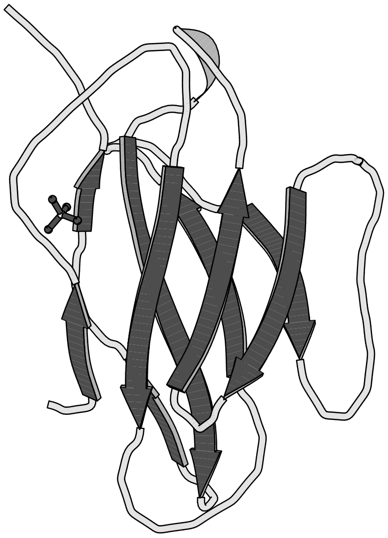
A large number of the all-β structures arrange their strands to form barrel-like or sandwich structure. The simplest of these arrangements is the up-and-down barrel or "clam motif" found in the retinol binding superfamily of proteins (Fig. 12a) where each strand adds to the next in an antiparallel manner until the barrel is complete. In this group of proteins the interior of the barrel provides a binding site for hydrophobic ligands.

A substantial number of all-β proteins are built from antiparallel strands in which adjacent strands are not directly connected. Many of these contain a topological feature known as a Greek Key in which the first strand connects across the top of the barrel or sandwich to the fourth



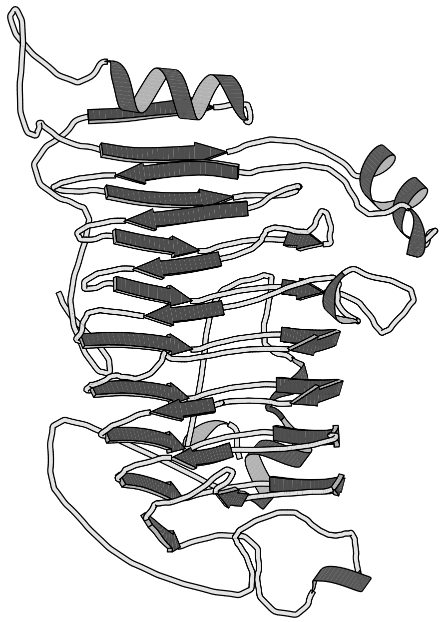
(a)

up-and-down barrel
Retinol binding Protein
(1RBP)



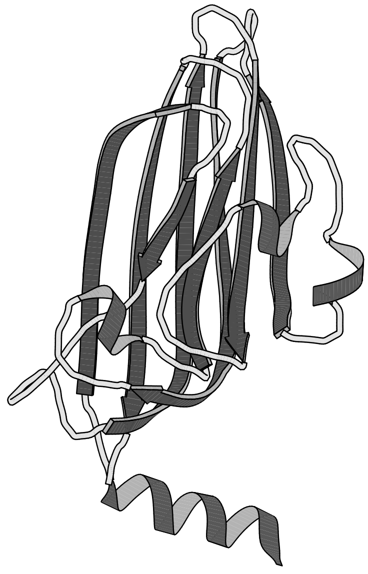
(b)

Cd8 Surface glycoprotein
N-terminal domain
Immunoglobulin fold
(Greek Key Motifs)
(1CD8)



(c)

Pectate lyase
(2PEC)



(d)

Sattelite Tobacco
necrosis virus
Jelly roll motif
(2STV)

FIGURE 12 Ribbon representations of typical all-β proteins. (a) Retinol binding protein, (b) immunoglobulin fold as seen in the Cd8 Surface glycoprotein N-terminal domain, (c) pectate lyase, and (d) viral coat protein found in satellite tobacco necrosis virus.

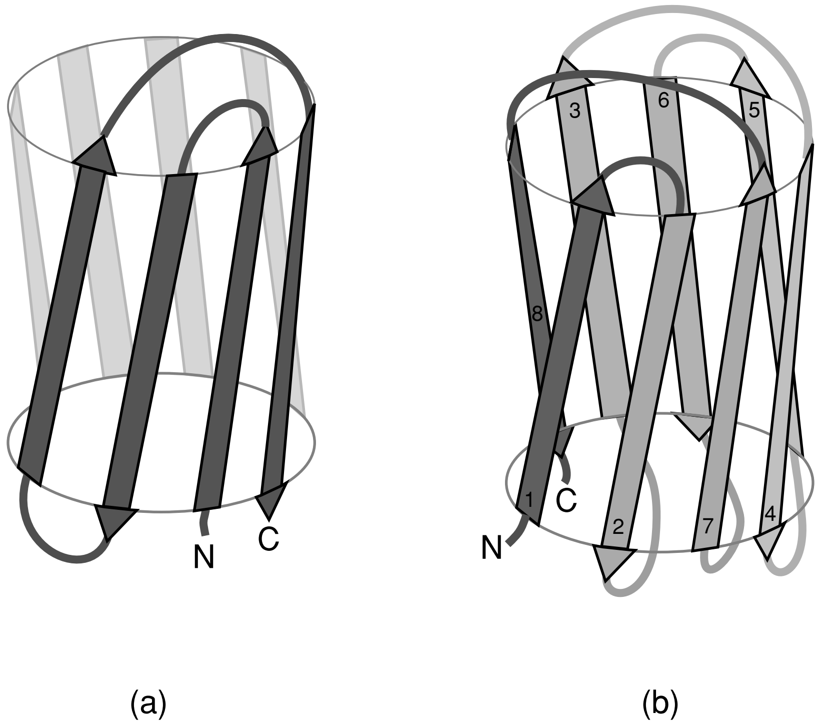


FIGURE 13 Schematic representations of (a) the "Greek Key" and (b) the "jelly roll" topologies commonly found in "all-β" proteins. These topological connections of strands have a handedness where only these arrangements are observed.

strand which then returns via two hairpin connections to the strand adjacent to the first. This feature contains a handedness that is only observed in one sense as shown in Fig. 13a. Connection of two of these features gives rise to an eight stranded β-barrel; however, Greek Key motifs are utilized in many ways to form closed structures. An alternative way of forming a closed barrel is found in proteins that exhibit a "jelly roll" topology (Fig. 13b). This is an abundant motif that is commonly found in virus capsid proteins.

The inclusion of a few sections of random coil or the occasional α-helix into an all-β protein allows for the generation of some remarkable motifs as shown in the β-propellors and β-helical folds. These folds illustrate the versatility of the β-strand when the hydrogen bonding potential of the polypeptide chain is fulfilled.

G. α/β Proteins

The α/β class of proteins contains many of the folds that incorporate parallel β-sheets. These folds exist in two major subclasses: the first contains a closed circular β-sheet surrounded by α-helices which forms a barrel; the second is based on an open sheet typically surrounded on both sides by α-helices (Fig. 14). Both of these arrangements are abundant in biosynthetic enzymes.

1. TIM Barrel

Triosephosphate isomerase was the first enzyme shown to contain an $(\alpha/\beta)_8$ barrel and thus established this motif as the TIM barrel. This fold consists of eight parallel β-strands connected by right-handed helical crossovers and is one of the most common folds found in enzymes. The TIM barrel typically contains approximately 200 amino acid residues. Contrary to the appearance of the ribbon drawing (Fig. 14a), the interior of the barrel is closely packed by the side chains protruding from the β-strands. The strands are inclined at an angle of approximately 30° to barrel axis, which is necessary to allow efficient packing of the interior. The necessity to form a closely packed interior explains why these barrels are almost always formed from eight strands. There are several variations on the TIM barrel that include the addition and subtraction of β-strands as well as the introduction of antiparallel β-strands as observed in enolase. These variations attest to the versatility of this fold.

The active sites of triosphosphate isomerase and all other enzymes that contain to this fold are located at the C-terminal end of the β-strands. Typically the catalytic residues reside at the end of the strands and are distributed around the barrel. The loops that connect the strands to the α-helices normally provide the components necessary for substrate specificity. The length of these connecting

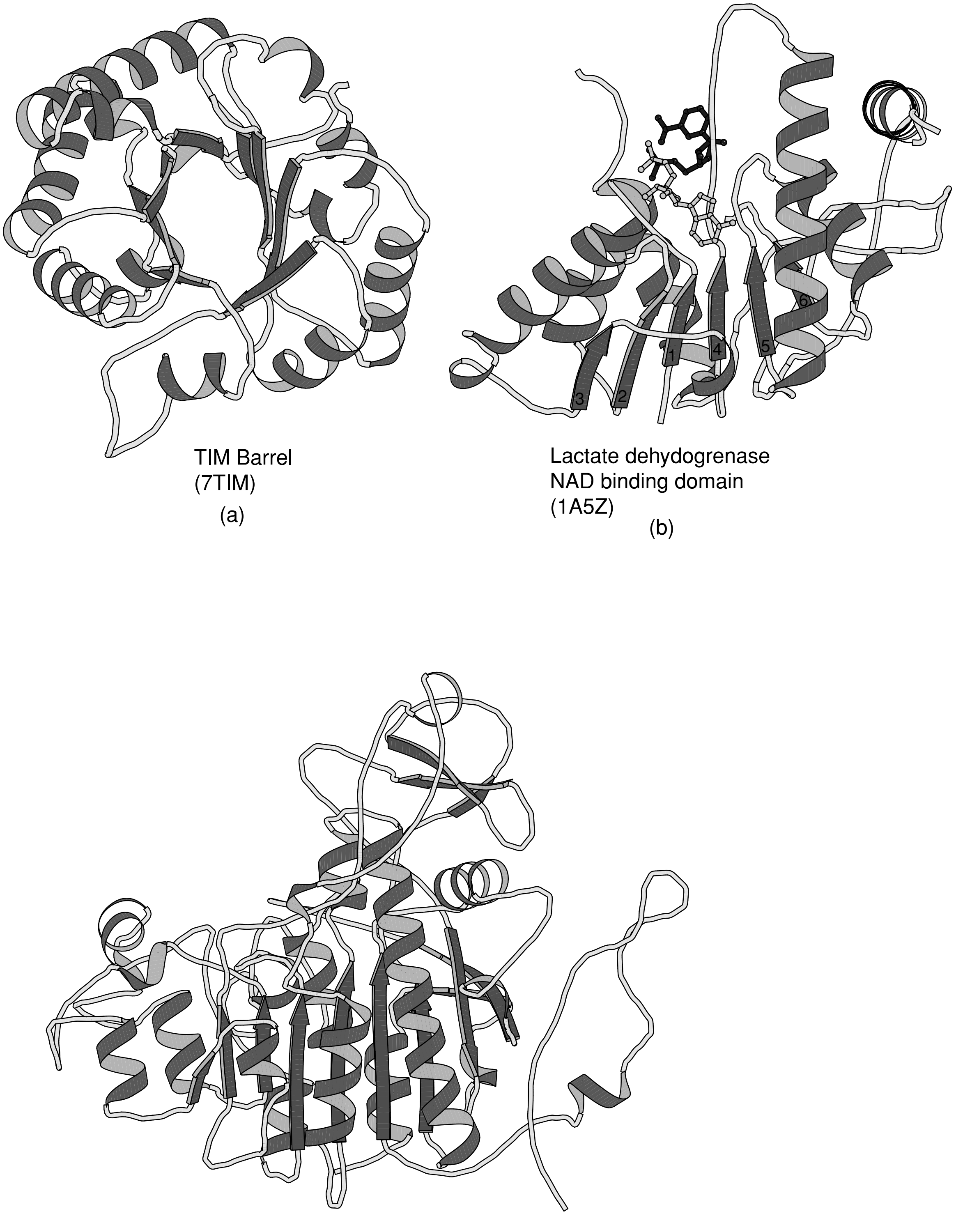


FIGURE 14 Ribbon representations of α/β proteins. (a) Triosephosphate isomerase, (b) dinucleotide binding domain of lactate dehydrogenase (c) alkaline phosphatase which is an example of a complex member of the α/β class of folds.

loops is enormously variable, whereas the length of the β-strands are similar in all enzymes.

2. Open β-Sheets

The second class of proteins in the α/β family of folds contains a large open sheet formed from mostly parallel β-strands with helices on both sides. In contrast to the TIM barrel there are fewer limitations on the number of strands within the sheet and may vary from 4 to 10. The first example of this type of fold was seen in lactate dehydrogenase which contains a motif that is widely observed in dinucleotide binding proteins (this motif is often referred to as the Rossmann fold) and was the first example of a domain superfamily (Fig. 14b). The observation of a common fold in the dehydrogenases by Rossmann and coworkers started the entire field of structural comparison and study of structural evolution.

All of the connections between β-strands are formed by right-handed crossovers. As a consequence, the strand order within the sheet must reverse in order to place helices on both sides of the sheet (Note: the consecutive strand order in the $(\alpha/\beta)_8$ barrel places the α-helices on one side of the sheet). In the classical Rossmann fold, which contains six β-strands, the N-terminal strand in the fold is located adjacent to the center of the motif. The first two α-helices lie on one side of the sheet as the first three strands are added. Thereafter the chain returns to the center of the sheet and adds the next three strands with the reverse strand order such that the subsequent helices are added on the opposite side of the sheet.

There are many varieties of open sheet α/β proteins which include differing numbers of strands, connections between strands that are not adjacent and incorporation of antiparallel strands. In most cases the ligand binding sites are located at the C-terminal ends of the β-strands and lie at the crevice at the edge of the sheet where the strand order is reversed. The loops that connect the strands to the helices typically provide the residues necessary for specificity. The size of the connecting loops are enormously variable in α/β proteins.

H. $\alpha + \beta$ Proteins

The $\alpha + \beta$ class of proteins is highly variable, indeed over a hundred distinct folds have been observed in this group. Members of this class typically contain one or more β-sheets which have a bias toward antiparallel connections. As such the α-helical and β-sheet regions tend to be segregated along their sequences. Several examples of proteins that fall in this class are shown in Fig. 15. In the simplest cases the helices lie on one side of the sheet which may be comparatively flat or steeply curved as in ubiquitin. These are known as $\alpha\beta$ folds. In more complex folds multiple layers of sheet and additional layers of helices have been observed to give rise to $\alpha\beta\beta$ (as in ribonuclease) and $\alpha\beta\beta\alpha$ folds (as in glutamine phosphribosyl pyrophosphate amidotransferase N-terminal domain).

I. Small Proteins, Unusual Folds

There are a substantial number of small proteins that defy classification into one of the groups listed above. Some of these have limited regular secondary structure whereas others are stabilized by metal ligands, cofactors, or disulfide bonds. Examples of these folds include the zinc–finger DNA binding motifs, many small iron–sulfur proteins, toxins and protein-inhibitors (Fig. 16).

VIII. MEMBRANE PROTEINS

Approximately one third of all proteins are tightly associated with membranes. These are much more difficult to crystallize or study by NMR than water-soluble proteins. As a consequence, there are far fewer structures of membrane proteins. Even so, those that have been determined provide insight into the manner in which polypeptide chains interact with lipid bilayers.

Membrane proteins fall into two classes: peripheral and integral. Peripheral membrane proteins are associated with the membrane, but may be removed by high concentrations of salt or metal chelators such as EDTA. In most aspects the structures of peripheral membrane proteins are very similar to water-soluble proteins. Integral membrane proteins differ in that they are very difficult to extract from the lipid bilayer and require detergents for solubilization. Detergents disrupt the lipid bilayer and bind to the hydrophobic surfaces of the protein that are buried within the membrane.

Integral membrane proteins all share the common problem of inserting a polypeptide chain into the hydrophobic interior of the lipid bilayer. This poses a thermodynamic problem on account of the hydrogen bonding propensity of the polypeptide chain. Clearly any segment of the protein that passes through the lipid bilayer must accommodate the hydrogen bonding potential of the polypeptide chain. Originally it was believed that an α-helix would be the only secondary structural element to pass through the lipid bilayer since it alone fulfills the hydrogen bonding capacity of the polypeptide chain in a consecutive manner. Indeed the α-helix is the only way to pass a single transmembrane segment of protein through a membrane. However, although a large number of membrane proteins are formed from α-helical bundles a significant number are built from β-strands. Both of these strategies for building

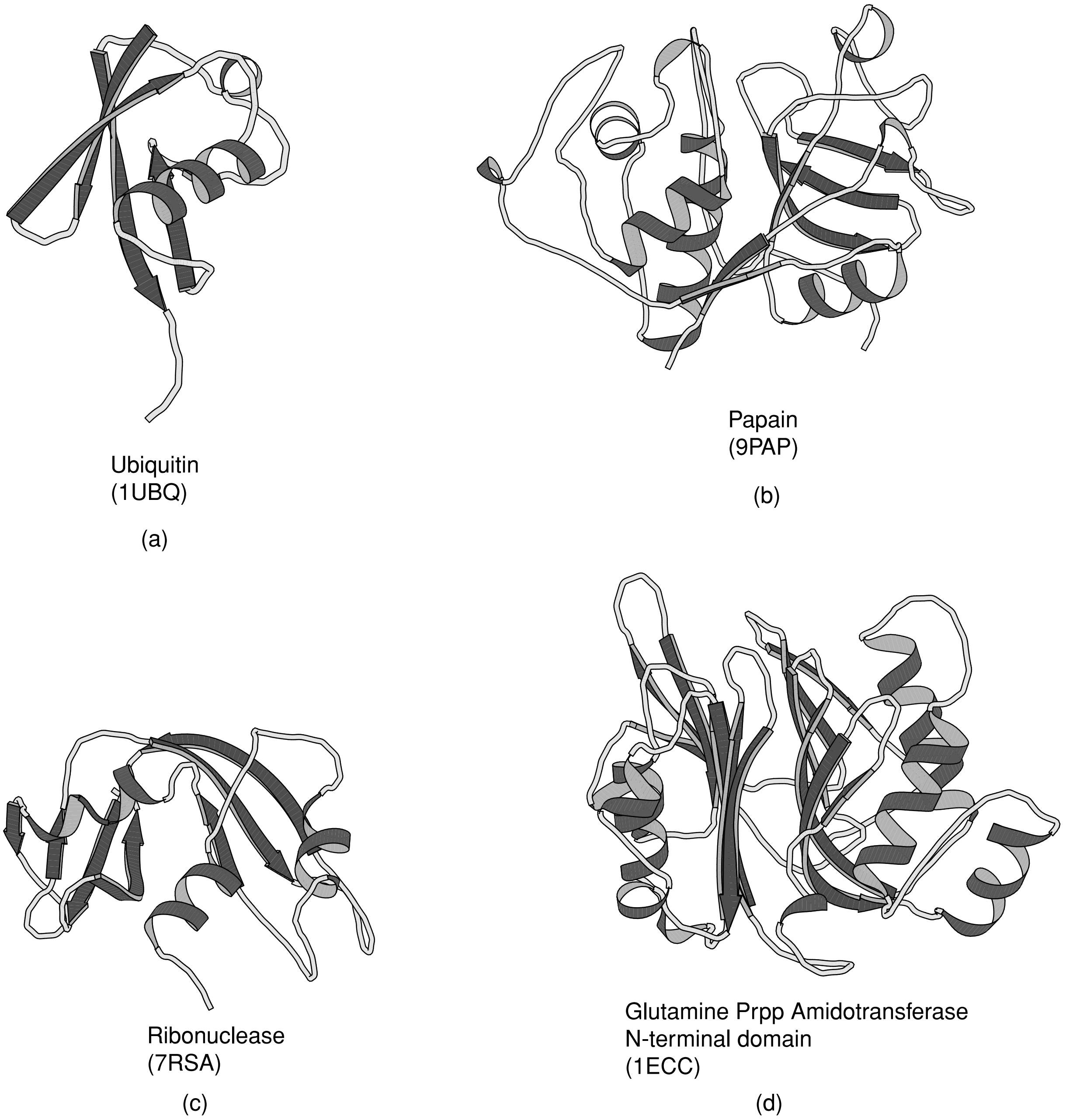
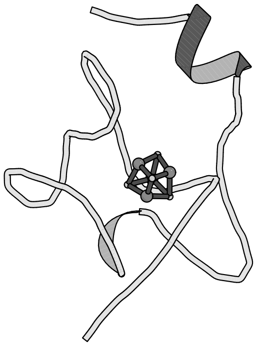


FIGURE 15 Examples of $\alpha + \beta$ proteins. (a) Ubiquitin, (b) Papain consists of one alpha-helix and four strands of antiparallel beta-sheet, (c) ribonuclease A, and (d) N-terminal domain of *E. coli* glutamine phosphoribosylpyrophosphate (Prpp) amidotransferase that contains a four structural layers; $\alpha\beta\beta\alpha$.

integral membrane proteins present interesting biophysical problems.

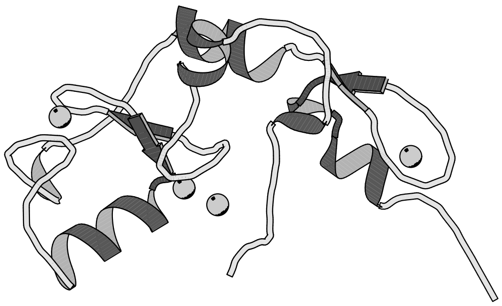
A. α-Helical Membrane Proteins

Many fully inserted membrane proteins utilize bundles of α-helices to span the lipid bilayer. The structure of the photoreaction center, the first membrane protein whose structure was determined, clearly shows this strategy (Fig. 17a). It consists of polypeptide chains that span the lipid bilayer utilizing 11 transmembrane α-helices. Interestingly the surfaces that face the interior of the bilayer are more hydrophobic than the interior of the protein, whereas the components that face the aqueous environments are similar to the surfaces of water soluble proteins. Thus there is no tendency for proteins to unfold in the lipid bilayer. This suggests that the same forces that stabilize water-soluble proteins are responsible for the stability of membrane proteins. α-Helices are the major component of proton



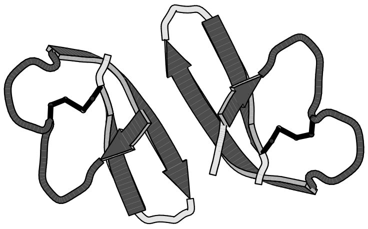
High potential Iron-Sulfur
Protein (1ISU)

(a)



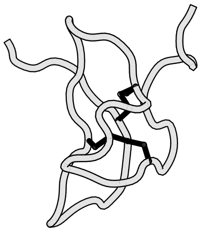
RAG1 DNA binding protein
Dimerization domain
(1RMD)

(b)



Defensin
amphiphilic dimer
(1DFN)

(c)



spider neurotoxin
(1QK6)

(d)

FIGURE 16 Ribbon representation of irregular structures: (a) High potential iron–sulfur protein coordinated to a 4Fe–4S cluster, (b) RAG1 DNA binding protein which contains representative examples of Zn–finger domains, (c) Defensin as example of a membrane toxin stabilized by disulfide bonds, and (d) Chinese bird spider neurotoxin which contains a cystine knot.

pumps such as bacteriorhodopsin (Fig. 17c) and K^+ channel proteins.

B. β-Sheet Membrane Proteins

The outer membranes of gram-negative bacteria contain channels that allow the diffusion of small solutes and ions into the periplasmic space. These channels are formed by bacterial porins that are built almost entirely of antiparallel transmembrane β-strands (Fig. 17b). The topology of these proteins is exceedingly simple, consisting of up-and-down strands where the first strand hydrogen bonds with the last strand of the sheet. In this way the hydrogen bonding potential of the β-strands is completely satisfied and a hydrophobic surface is presented by the side chains that extend into the lipid bilayer. The channel lies down the middle of the barrel. The side chains that line the pore provide some selectivity for the nature of the solutes that diffuse through the channel. All bacterial porins exist as oligomers (mostly trimers) where the interface between the porin subunits form a hydrophobic interior that is otherwise missing from these proteins.

This type of antiparallel packing of β-strands in integral membrane proteins has also been observed in hemolysin. Hemolysin is a heptameric pore forming protein from *Staphylococcus aureus*, where each of the subunits contributes two antiparallel β-strands to the transmembrane segment creating a 14 stranded barrel.

C. Other Membrane Motifs

Not all membrane proteins extend completely through the lipid bilayer, indeed many of the biosynthetic enzymes that are tightly bound to the membrane only extend into

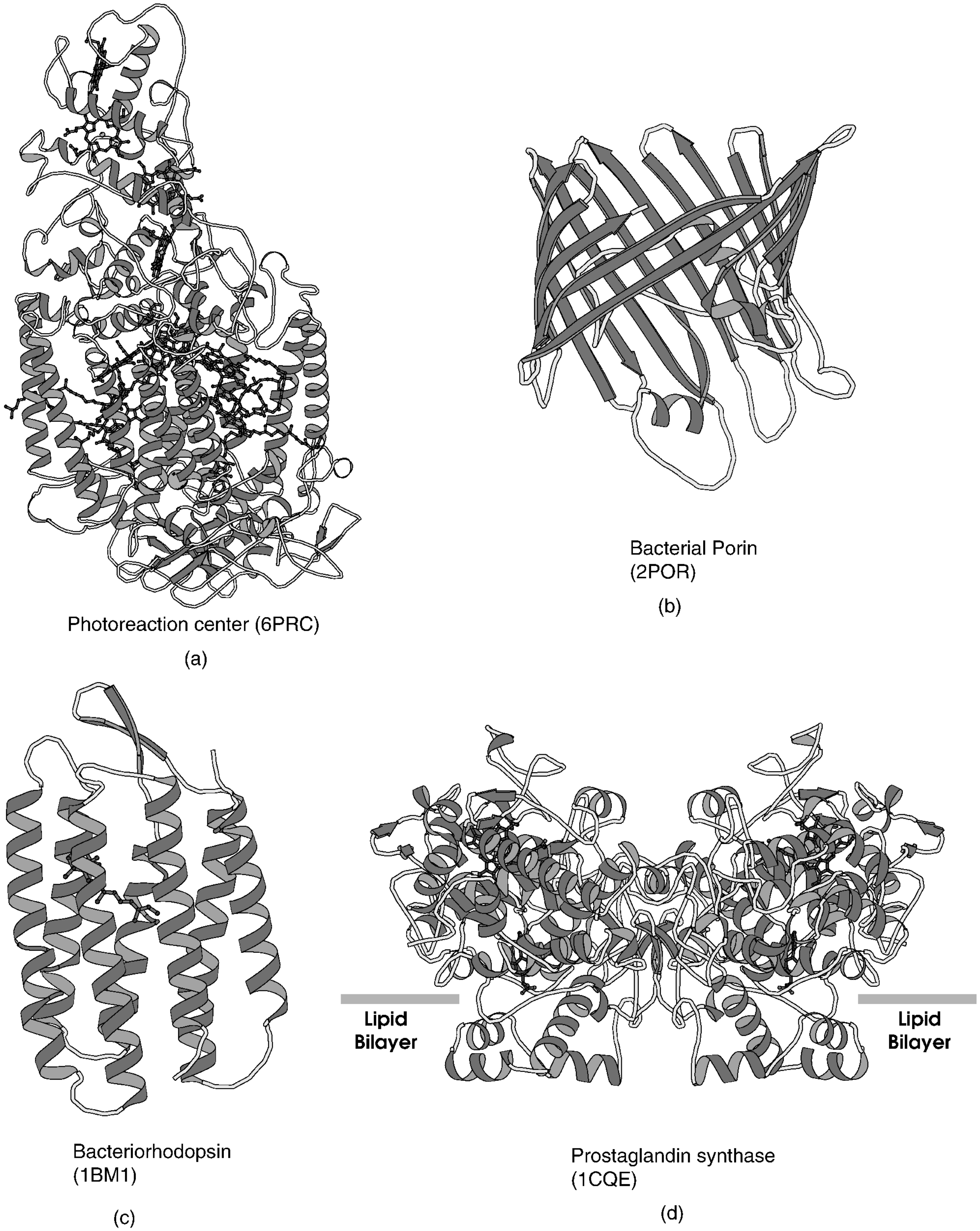


FIGURE 17 Ribbon representation of selected integral membrane proteins. (a) Photoreaction center, (b) bacterial porin, (c) bacteriorhodopsin, (d) prostaglandin synthase.

one face. A good example of this is prostaglandin synthase (Fig. 17d). This enzyme converts arachidonic acid into prostaglandin (PGH_2) and exhibits two catalytic activities: a cycloxygenase and peroxidase. It is an important enzyme since it catalyzes a critical step in the biosynthesis of a wide range of eicosanoids that control the inflammatory response, pain and fever, blood pressure, induction of blood clotting, induction of labor, and the sleep/wake cycle. In most respects the structure of this dimeric enzyme is very similar to other water enzymes, except that it exhibits a highly hydrophobic surface that is built from four α-helices per subunit that are inserted into the lipid bilayer. The use of α-helices to penetrate one side of a lipid bilayer would appear to be a common feature of proteins that interact with one side of a membrane. Interestingly the active site opens to the lipid bilayer via a large tunnel. This provides a thermodynamically acceptable route for arachidonic acid, which is a hydrophobic substrate to enter the enzyme. Prostaglandin synthase is the site of action of nonsteroidal anti-inflammatory drugs such as aspirin and ibuprofen, which act to block the channel that opens to the lipid bilayer and prevent arachidonic acid from being converted to PGH_2.

IX. SUPERFAMILIES AND STRUCTURAL EVOLUTION

The topology of a domain yields information about its evolutionary history. Extensive studies on the sequence variation among a family of similar enzymes found in differing organisms reveal that for a given biological function the protein fold is more conserved than the sequence, except for catalytically vital residues. Remarkably the same protein fold is found in proteins that share no significant sequence similarity. Together these observations suggest that a given fold can accept a wide range of sequences and that evolution preserves the core three-dimensional structure. There are two explanations for these observations. The first explanation suggests that evolution of proteins occurs primarily through point mutations which requires that the evolutionary intermediates must be functional and stable. Clearly this is required if the changes involve an essential enzyme. The same requirement applies to more drastic genetic rearrangements which must also proceed through useful, stable intermediates to survive selective pressure and again this will preserve the protein fold. An alternative explanation is that there are a limited number of stable folds and that enzymes have evolved to reach these conformers. In all likelihood both of these arguments contain elements of truth.

There is no doubt that nature has frequently adapted successful protein architectures to carry out new biological functions. This can be seen clearly in the repeated use of common ligand binding domains (such as the dinucleotide binding motif) that occur repeatedly in proteins that require the same ligand even it is used in differing chemical reactions. Such use of common building blocks can be rationalized as a consequence of genetic rearrangement. It would appear that some folds, such as the TIM barrel, are particularly well suited for the evolution of new functionalities. The TIM barrel is one of the most abundant enzyme folds and appears to have arisen at least twice during evolution on the basis of the hydrogen bonding pattern observed within the barrel. There are a wide number of enzymes that utilize this fold as the foundation of their active sites which suggests that enzymes may evolve by retooling of existing functional folds.

The enolase superfamily, which contains a variant of the TIM barrel, is a good example of retooling of a functional fold since these enzymes share a common catalytic step of abstraction of the α-proton of a carboxylate anion. This group of enzymes catalyzes a remarkable range of chemical reactions including racemization, β-elimination of water, β-elimination of ammonia, and cycloisomerization. Each enzyme contains similar catalytic bases and acids and each appears to have evolved by reusing a structural framework that facilitates a difficult chemical task. This is consistent with the observation that the protein fold is the component of the structure that changes most slowly during evolution. Fundamentally this is the result of the marginal thermodynamic stablity of proteins.

X. QUATERNARY STRUCTURE

Although a substantial number of proteins function as monomers there are many others that exist as multimers. The arrangement of protein subunits in a macromolecular assembly is referred to as its quaternary structure. This aspect of protein structure plays an important role in the stability and regulation of a large number of enzymes, virus assembly, cellular regulation, and motility. Indeed, quaternary structure underlies all aspects of protein–protein interaction.

The magnitude of protein–protein interfaces varies enormously from $\sim$800 Å^2 to over 4000 Å^2 per subunit. There is considerable variation in the nature of the interface. Some interfaces are very similar to the hydrophobic core the protein, whereas others contain a substantial polar component and solvent pockets. Thus the stability of a multimeric assembly is only loosely proportional to the surface area buried on formation of the multimeric assembly. As might be expected from considerations of

protein stability, the overall stability of multimeric assemblies is proportional to the hydrophobic surface area buried within the interface.

A large number of enzymes exist as symmetric macromolecular assemblies where they commonly exhibit C and D point group symmetries and contain two-, three-, four-, and six-fold axes of symmetry. In the simplest cases this feature provides additional thermodynamic stability for a protein that would otherwise be rather small. In more complex arrangements the protein–protein interfaces form the active site such that the oligomerization is required for function. Finally, in the most highly evolved enzymes there is communication between the active sites that reside on symmetrically related subunits. This provides the foundation for enzyme regulation as observed in most allosteric enzymes.

A. Allosteric Enzymes

The simplest model for allosteric control was set forward by Monod, Changeux, and Jacob in 1963 and provided the basis for understanding feedback inhibition and cooperative binding of ligands by proteins. In their model it was assumed that an allosteric enzyme (or protein) exists in equilibrium between two symmetric states; inactive and active (T and R states). The transition between these states was assumed to be concerted, that is the symmetry of the macromolecular assembly is conserved. Furthermore the active state has a greater affinity for substrate than the inactive state. From these considerations the activity of the enzyme depends on the position of the equilibrium between the inactive and active states. Thus increasing the substrate concentration drives the equilibrium to the active form and gives rise to a sigmoidal relationship between the initial velocity of the reaction and the substrate concentration. As importantly the position of the equilibrium can be altered by allosteric effectors that preferentially bind to either the inactive or active state. This is the basis of feedback inhibition whereby the product of a biosynthetic pathway inhibits the enzyme that catalyzes the first committed step. This simple model explains many of the properties of allosteric enzymes; however, a more complex model based on sequential binding of substrates to yield multiple conformation states is required to explain the finer details of many enzymes.

The structural basis of allostery has been well developed through the study of enzymes such as aspartate transcarbamoylase and phosphofructokinase. In all enzymes studied thus far several common themes have evolved. First the overall symmetry of the inactive and active states appear to be conserved. Second the transition between these states involves a change in the relationship between the protein subunits that serve to alter the position of the polypeptide chains or cofactors within the regions responsible for ligand binding or catalytic activity. Allosteric effectors serve to either inhibit or enhance these conformational changes. Thus in all cases the active site residues are influenced either directly or indirectly by the protein–protein interactions between subunits. Typically a rotation of subunits of 5–15° is sufficient to accomplish allosteric control.

B. Viruses

Viruses represent particularly evolved forms of macromolecular assemblies. Mature virions are encoded by a protective coat that is formed in part by virally encoded proteins. Viruses come in many shapes and sizes, but they all share the property of using multiple copies of coat proteins to protect their genomic material. In many cases the proteins assemble to form a symmetric shell, where the symmetries are either helical as found in tobacco mosaic virus or icosahedral as seen in the spherical viruses. The use of multiple copies of a protein to form a viral coat is enormously efficient from a genomic point of view; however, it introduces several interesting structural problems.

All simple spherical viruses exhibit icosahedral symmetry. This implies that the surface contains 60 equivalent positions or that the shell is built from 60 equivalent protein subunits. Although this is observed for a few very small virus particles, most contain many more than 60 subunits. For example the viral shell of most small plant viruses contains 180 identical protein subunits. This poses a problem of how to arrange 180 subunits on the surface of an icosahedral shell since the subunits cannot experience symmetrically equivalent environments. Fifty years ago it was proposed by Caspar and Klug that viral subunits would be arranged on a hexagonal surface lattice with quasiequivalent symmetry where the contacts between subunits would be organized to minimize their differences in assembly. That is, if a virus contains 180 subunits on its surface, these would be grouped into three sets of 60 subunits where each group would experience similar interactions compared to the other groups. At first sight this hypothesis accounts for the structure of simple viruses where it has been shown that most of the interactions between the protomers on the surface are essentially identical; however, closer examination of the virus structures that have been determined reveals that in all cases the groups of protomers behave as though they are different proteins by utilizing their domains in different ways to accommodate their structurally unique environments (Fig. 18). Thus, it would appear that the main requirement for a virus coat protein is to have a shape and conformational flexibility to

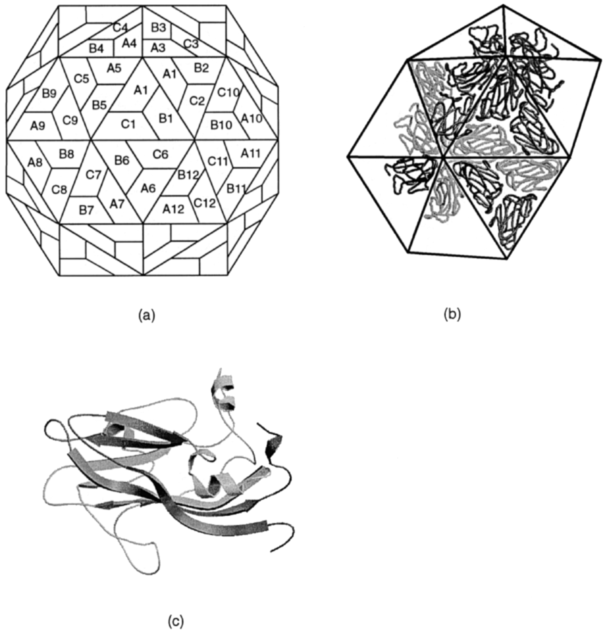


FIGURE 18 Structural arrangement of capsid proteins in southern bean mosaic virus. (a) The surface lattice describing the relationship between the three groups of 60 identical subunits. (b) The arrangement of protein subunits on the surface of the virus. (c) A ribbon representation of viral coat protein for position A.

allow them to form a cohesive shell, without compromising the structural integrity of the quaternary interactions. These latter principles may be applied to all viral coat proteins.

XI. CONCLUSIONS

Protein structure influences all aspects biological function. Although there is considerable variation in the structural motifs observed in biological macromolecules they are all unified by being built from the same 20 amino acids. The differences are due to the essentially infinite number of protein sequences that may be generated from these building blocks. At a fundamental level, rigorous knowledge of the conformational properties of polypeptide chains should lead to a complete understanding of protein structure and function. As described earlier, great progress has been made toward this goal, but much remains to be done.

SEE ALSO THE FOLLOWING ARTICLES

Bioenergetics • Hydrogen Bond • Macromolecules, Structure • Membrane Structure • Nuclear Magnetic Resonance (NMR) • Protein Folding • Protein Synthesis • X-Ray Analysis

BIBLIOGRAPHY

Branden, C., and Tooze, J. (1999). "Introduction to protein structure," (2nd edition), Garland.

Carter, C. W., Jr., Sweet, R. M., Abelson, J. N., and Simon, M. I. (eds.) (1997). "Methods in Enzymology," Volume 276, Macromolecular Crystallography, Part A, Academic Press, New York.

Carter, C. W., Jr., Sweet, R. M., Abelson, J. N., and Simon, M. I. (eds.) (1997). "Methods in Enzymology," Volume 277, Macromolecular Crystallography, Part B, Academic Press, New York.

Cavanagh, J., Palmer, A. G., III, Fairbrother, W., and Skelton N. (1996). "Protein Nmr Spectroscopy: Principles and Practice," Academic Press, New York.

WEB SITES

Data Bases for Macromolecular Structures

The RSCB (//www.rcsb.org/pdb/). The most comprehensive archive for coordinates of proteins and nucleic acids (PDB files).

//www.bmrb.wisc.edu/pages/ A repository for data from NMR spectroscopy on proteins, peptides and nucleic acids

Sites for Structural Analysis

CATH (//HYPERLINK http://www.biochem.ucl.ac.uk/bsm/cath/www.biochem.ucl.ac.uk/bsm/cath/) classification, and

SCOP (//scop.mrc-lmb.cam.ac.uk/scop/) structural databases.

//mmtsb.scripps.edu/viper/viper.html. Data base for virus structural information

//www.biochem.ucl.ac.uk/bsm/pdbsum. A site that summarizes structural analysis of PDB files. This contains links to the major structural classification sites.

//www2.ebi.ac.uk/dali: The Dali server is a network service for comparing protein structures in 3D. Coordinates of a query protein structure are submitted and Dali compares them against those in the Protein Data Bank. A multiple alignment of structural neighbours is returned.

Protein Synthesis

Paul Schimmel
Scripps Research Institute

Rebecca W. Alexander
Wake Forest University

I. Information Transfer and the Genetic Code
II. Transfer RNAs
III. Aminoacyl–tRNA Synthetases
IV. An Overview of Translation
V. Translation Initiation
VI. Elongation
VII. Termination
VIII. Ribosome Recycling
IX. Molecular Mimicry by Translational Factors
X. Translational Accuracy
XI. The Ribosome
XII. Inhibitors of Protein Synthesis
XIII. Post-Translational Modifications
XIV. Mitochondrial Protein Synthesis
XV. Nonribosomal Peptide Synthesis
XVI. Protein Synthesis and the RNA World

GLOSSARY

Aminoacylation Attachment of an amino acid to its cognate (matching) transfer RNA (tRNA), catalyzed by the cognate aminoacyl–tRNA synthetase (AARS).

Anticodon The trinucleotide sequence at the end of one arm of tRNA that base pairs with a complementary messenger RNA (mRNA) codon.

Codon The trinucleotide sequence of an mRNA that specifies which amino acid will be inserted into a protein.

Genetic code The set of rules that specify codon-amino acid correspondence.

Translation Synthesis of a protein according to its mRNA sequence.

Watson–Crick base pair Hydrogen bond partners in RNA or DNA. Adenine is the Watson–Crick partner to uracil (or thymine in DNA); guanine pairs with cytosine.

PROTEINS are polymers of amino acids joined by peptide bonds (proteins are therefore also known as polypeptides). The number and order of the amino acids contained in a particular protein are prescribed by the DNA sequence of that protein's gene. The mechanism by which a genetic message is translated from its nucleic acid form to its polypeptide product is protein synthesis. Translation occurs at the ribosome, an RNA–protein complex often called the protein factory of the cell.

I. INFORMATION TRANSFER AND THE GENETIC CODE

Many cellular components play a role in protein synthesis (Fig. 1). Even in relatively simple bacteria, translation of a single polypeptide from its genetic message requires dozens of participants—proteins, RNAs, and nucleotides—working together as carriers, catalysts, energy

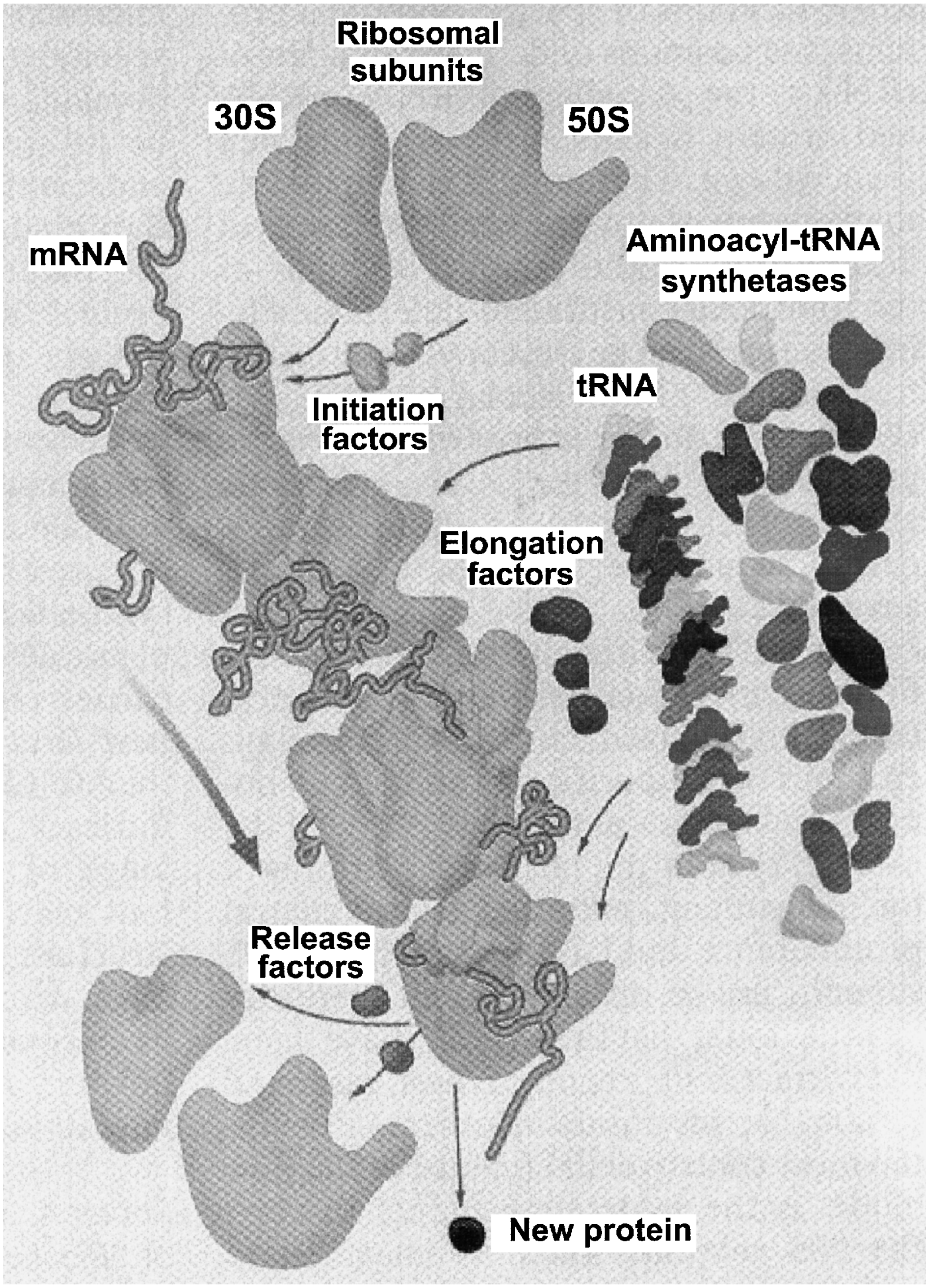


FIGURE 1 Bacterial protein synthesis. This cartoon summarizes the numerous components necessary for translation of an mRNA into its corresponding polypeptide product. Translation occurs at the ribosome, and is initiated at an AUG start codon by protein initiation factors. Aminoacyl–tRNA synthetases attach amino acids to tRNA adaptor molecules in highly specific enzymatic reactions. Each elongation cycle adds an amino acid to the growing polypeptide chain according to base-pairing interactions between each trinucleotide codon of the mRNA and the anticodon of the matching aminoacylated tRNA. Elongation factors facilitate aminoacyl tRNA selection and, following peptide bond formation, translocation of the peptidyl–tRNA and mRNA. When a stop codon is reached, release factors trigger hydrolysis of the newly synthesized protein and dissociation of ribosomal subunits. [From Zimmermann, R. A. (1995). "Protein synthesis. Ins and outs of the ribosome." *Nature* **376,** 391–392. © 1995 Macmillan Magazines Ltd.]

sources, and cofactors. Peptide bond formation takes place rapidly at the ribosome, with as many as 40 amino acids per second joined to a growing polypeptide chain. Yet coupled with the need for speed is the requirement for accuracy. The misincorporation of a single amino acid could have drastic effects on the structure or function of a newly synthesized protein. However, protein synthesis is accurate, with errors occurring on the average only once in 10,000 peptide bonds formed.

Formation of peptide bonds linking together amino acids could theoretically occur such that random sequences are generated. Some of these sequences could result in a polypeptide that has a useful function. However, transfer of genetic information from one generation to the next requires a systematic and reproducible mechanism for generating defined sequences. Polypeptide formation as we know it today is template-directed, with the messenger RNA (mRNA) copy of a gene providing the text to be deciphered into the protein product.

The simplest link between nucleic acid and protein components would have been a code with a one-to-one correspondence where each nucleotide dictated a particular amino acid. With only four nucleotides making up the information storage in cells, the resulting proteins synthesized in such a scenario would be limited to those having 4 different amino acids. Even a code of two nucleotides per amino acid would allow for only 16 amino acids. The standard genetic code instead makes use of trinucleotide sequences called codons; these 64 codons are able to determine fully the 20 amino acids used in protein synthesis and also include start and stop codons (Table I).

TABLE I The Standard Genetic Code

First position	Second position U	Second position C	Second position A	Second position G	Third position
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met[a]	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

[a] The AUG codon specifies the start of protein synthesis as well as internal methionine residues.

II. TRANSFER RNAs

Although an mRNA nucleotide sequence dictates the polypeptide sequence to be made, mRNAs do not directly recognize amino acids. Amino acids are instead linked to transfer RNA (tRNA) "adaptor" molecules, which serve as reading heads to decipher the codons of mRNA through base-pairing complementarity (Fig. 2).

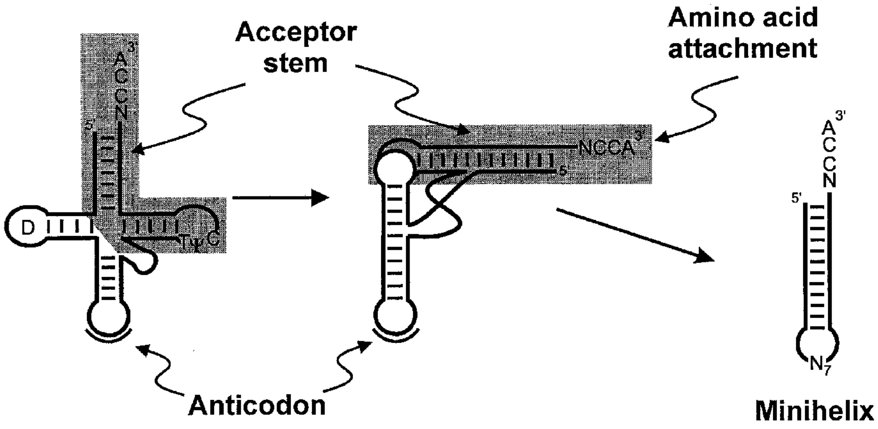


FIGURE 2 Transfer RNA folding. The tRNA cloverleaf secondary structure representation (left) folds into an L-shaped structure consisting of two domains (center). The highly conserved D and TΨC loops are indicated. Tertiary interactions, including unusual base pairs and base triples facilitate and stabilize formation of the corner of the L. The anticodon is separated approximately 75 Å from the site of amino acid attachment. The RNA minihelix domain is highlighted by shading, while the second domain containing the anticodon and D-arm is unshaded. Minihelix RNAs are in many systems substrates for aminoacylation by aminoacyl-tRNA synthetases.

A. Conserved Features of tRNAs

Transfer RNAs typically contain approximately 76 nucleotides and have a molecular mass of about 25 kD. The characteristic "cloverleaf" secondary structure representation of tRNA was predicted based on regions of base complementarity, and the determination of hundreds of tRNA sequences demonstrated that such a folding pattern is conserved. Several sequence and structural features are also consistently present in tRNAs from all organisms. The sequence at the 3′-end of tRNAs is always –CCA, with a free hydroxyl group on the terminal adenosine that is the site of amino acid attachment. Nucleotides near the termini hybridize to make the 7-base-pair (bp) acceptor stem.

The other arms of the tRNA cloverleaf also have distinctive conserved features. The modified base dihydrouridine (D) is typically present in the loop that closes off a short 3- or 4-bp stem following the acceptor stem. This stem and loop are therefore called the D-arm. The anticodon arm consists of a 5-bp helix closed by a loop that contains the trinucleotide anticodon. Following the anticodon arm is the variable loop, which can contain 3–21 nucleotides, with a stem as long as 7 bp, depending on the particular tRNA. The modified bases pseudouridine (Ψ) and ribothymidine (T) are usually present in the loop of the TΨC arm, so named because of the presence of this highly conserved sequence.

B. The L-Shaped Structure of tRNAs

In three dimensions, tRNAs fold into an L-shaped structure in which the acceptor stem and TΨC arm coaxially stack to form one part of the L known as the minihelix, and the D and anticodon arms likewise stack to form the other part of the molecule. This structure is facilitated and stabilized by tertiary interactions at the corner of the L that bring together the D and variable loops. The nucleotides involved in these interactions are typically invariant or semi-invariant, indicating that the tRNA L shape is universal. While most base pairs in tRNA helices are canonical Watson–Crick pairs, the tertiary interactions at the corner of the L make use of some unusual hydrogen-bonding conformations. For example, nearly all tRNAs contain a U8:A14 reverse Hoogsteen base pair, and several base triples (where three bases are paired together) are also typically present at the core of the structure.

The two portions of the L can be considered distinct domains with separate contributions to protein synthesis. The minihelix containing the acceptor arm includes the site of amino acid attachment (the 3′-OH). It is considered by many investigators to be related to the historical or early form of tRNA. The anticodon trinucleotide is located at the other end of the L, approximately 75 Å away. This separation between anticodon and amino acid on the tRNA is paralleled in the mRNA decoding and peptide bond formation events that occur on two different subunits of the ribosome. Furthermore, the domains of tRNA can be physically separated such that an isolated acceptor stem ("minihelix") can in many cases accept its specified amino acid, while an anticodon stem-loop helix can bind to the ribosome-bound mRNA.

C. Codon–Anticodon Interactions

Translation of a genetic message into its protein product depends on base pairing interactions between mRNA codon and tRNA anticodon. The 64 trinucleotide codons are more than sufficient to fully determine all 20 amino acids as well as one start and three stop codons. Thus, the code is degenerate, with many amino acids having more than one codon. This codon degeneracy is primarily due to variation in the third position of the trinucleotide, as shown in Table I. Although many organisms have more than one tRNA molecule per amino acid (these are "isoacceptors"), in many cases a specific tRNA recognizes more than one codon. Non-Watson–Crick base pairs are permitted at the third position, because of room for some structural flexibility or "wobble" in the pairing geometry. For example, a U in the third position (5′ nucleotide) of the anticodon can base pair with an A (Watson–Crick pair) or a G (wobble pair) in the codon. Several tRNAs contain an inosine (I) nucleotide at the third anticodon position; inosine forms a standard base pair with C and wobble pairs with U and A.

III. AMINOACYL–tRNA SYNTHETASES

Decoding of the protein message occurs at the ribosome, after prior attachment of amino acids to the tRNA molecules. Aminoacyl–tRNA synthetases (AARSs) are the family of enzymes responsible for covalent attachment of each amino acid to its correct, or cognate, tRNA molecule. This first step in protein synthesis is responsible for establishing the rules of the genetic code (which codon corresponds to which amino acid), as the aminoacylated tRNA has both the nucleic acid component of the genetic message (the anticodon) and the amino acid. In most organisms, there are 20 AARSs, one for each amino acid. Because of codon degeneracy, there is at least one tRNA per amino acid; serine, for example, can be attached to one of five different isoaccepting tRNA molecules. Although each enzyme catalyzes the same aminoacylation reaction, the substrates are unique and must be effectively selected from among the cellular pool. For example, valyl–tRNA synthetase (ValRS) binds the amino acid valine

specifically and disregards the other 19 amino acids; it also recognizes tRNAs containing the anticodons GUN (where N is any nucleotide) and only attaches valine to these molecules.

A. Amino Acid Attachment

Covalent attachment of an amino acid to its cognate tRNA occurs in a two-step mechanism catalyzed by an AARS (E). In the first step, energy is consumed in the form of ATP to convert the enzyme-bound amino acid to its high-energy aminoacyl adenylate (AA–AMP). In the second step the activated amino acid is transferred to the 3′ end of the cognate tRNA, where an ester bond is formed between the carboxyl group of the amino acid and either the 2′- or 3′-hydroxyl on the terminal adenosine's ribose sugar to yield AA–tRNA. While for most AARSs the first step is tRNA-independent, glutaminyl–, glutamyl–, and arginyl–tRNA synthetases require cognate tRNA binding for aminoacyl adenylate formation.

$$\mathrm{E + AA + ATP \rightleftharpoons E(AA\text{–}AMP) + PP_i}$$

$$\mathrm{E(AA\text{–}AMP) + tRNA \rightleftharpoons AA\text{–}tRNA + E + AMP.}$$

B. Selection of tRNA

Because of the critical need for accurate tRNA aminoacylation in protein synthesis, AARSs must be highly discriminating in substrate selection. Structural studies revealed a large area of contact between AARS and tRNA, providing numerous opportunities for sequence-specific recognition, including contacts with the many modified nucleotides of tRNAs. The majority of identity determinants lie in the anticodon and the acceptor stem (near the site of amino acid attachment) of the tRNA molecule. Although the genetic code links amino acid identity to anticodon sequence, an "operational RNA code" exists within the acceptor stems of tRNAs. This operational RNA code is considered by some to be a "second genetic code." It relates sequence and structure in the acceptor stem to aminoacylation with specific amino acids.

Because determinants for aminoacylation are found in the acceptor stem, RNA minihelices are substrates for aminoacylation by at least 10 different aminoacyl–tRNA synthetases. A striking example is alanyl–tRNA synthetase, which makes no contact with the $tRNA^{Ala}$ anticodon. Instead, the enzyme recognizes a unique G3:U70 base pair in the acceptor stem of the tRNA. Substitutions at this location eliminate aminoacylation of the tRNA with alanine, while introduction of this base pair into noncognate tRNA molecules target them for alanylation. An RNA "minihelix" (Fig. 2) is also efficiently aminoacylated with alanine provided it contains the critical G3:U70 pair. Thus, selection of the cognate tRNA by an AARS is based on a limited number of sequence or structural elements, and in some cases does not involve the anticodon triplet of the genetic code.

C. Amino Acid Selection

Because of their small size and limited chemical diversity, selection of a cognate amino acid from the cellular pool presents a greater challenge to the AARSs than does selection of a specific tRNA. Differentiation of amino acids is primarily achieved through the formation of a binding pocket specific for the cognate amino acid within the active site of the enzyme. In this way specific hydrogen bond contacts can be made to a polar amino acid, or a positively charged amino acid can fit into a negatively charged cleft within the AARS. Amino acids with hydrophobic side chains have to be discriminated on the basis of shape.

In some cases binding of cognate tRNA influences amino acid selection. Glutaminyl–tRNA synthetase (GlnRS) is one of three AARSs that require the presence of cognate tRNA for aminoacyl adenylate formation. Biochemical and genetic studies demonstrated that conserved residues of GlnRS that contact the 3′-terminus of $tRNA^{Gln}$ are required to properly form the glutamine binding site. The presence of noncognate tRNAs decreased the enzyme's affinity for glutamine because only $tRNA^{Gln}$ was properly oriented to allow formation of the glutamine binding pocket.

D. Editing by AARSs

Binding interactions alone can not completely discriminate against all noncognate amino acids. Some enzymes misactivate noncognate amino acids at a significant frequency and require an additional accuracy mechanism. The most well-characterized example is isoleucyl–tRNA synthetase (IleRS), responsible for activating isoleucine and attaching it to $tRNA^{Ile}$ at the exclusion of valine, which is smaller than isoleucine by a single methylene group. IleRS is more successful in discriminating against larger or bulkier amino acids, but the smaller valine simply fits more loosely in the enzyme active site.

Despite misactivating valine as frequently as once per 180 correct (isoleucine) activations, the enzyme is able to maintain the required fidelity of tRNA aminoacylation through an active editing function, so that valine is not misincorporated into proteins to a significant extent. This editing is twofold, corresponding to the two steps of the aminoacylation reaction: pretransfer editing by IleRS selectively hydrolyzes the noncognate valyl–adenylate, while post-transfer editing hydrolyzes any valyl–$tRNA^{Ile}$ that is formed.

$$\text{IleRS : Val-AMP} + \text{tRNA}^{\text{Ile}} \rightarrow \text{IleRS} + \text{Val} + \text{AMP} + \text{tRNA}^{\text{Ile}} \text{ (Pretransfer)}$$

$$\text{IleRS : Val-tRNA}^{\text{Ile}} \rightarrow \text{IleRS} + \text{Val} + \text{tRNA}^{\text{Ile}} \text{ (Post-transfer).}$$

Both activities are dependent on the presence of cognate $tRNA^{Ile}$, particularly the D-arm, as determined by studies that investigated the ability of mutants of $tRNA^{Ile}$ to trigger editing. Furthermore, the editing site on IleRS has been located within a peptide (connective polypeptide 1, CP1) inserted between the two halves of the Rossmann nucleotide binding fold that makes up the catalytic domain of Class I AARSs (see following). Valyl–tRNA synthetase (ValRS) and leucyl–tRNA synthetase (LeuRS) have sequences in CP1 similar to IleRS; it is interesting to note that these enzymes all recognize amino acids with aliphatic side chains. Isolated CP1 domains from IleRS and ValRS have been shown to execute post-transfer editing.

The location of the editing site in IleRS has been visualized by X-ray crystallography and is approximately 25 Å from the "synthetic" active site. Apparently both the aminoacyl adenylate and the aminoacylated tRNA must be translocated from the synthetic active site to the editing active site for an accuracy check before being released. Translocation of the misaminoacylated tRNA to the editing site, and subsequent proofreading, has been proposed to be triggered by a conformational change in the tRNA. How the isolated aminoacyl adenylate is translocated remains unknown. The overall mechanism for accurate aminoacylation by IleRS has been termed a "double-sieve"—amino acids that don't fit into the active site binding pocket are excluded based on size, while valine and smaller amino acids are actively hydrolyzed at the second (editing) site when they are misactivated or subsequently attached to $tRNA^{Ile}$.

E. Class Organization of AARSs

Detailed structural information from X-ray crystallographic studies is available for nearly all of the 20 AARSs. As more and more sequences and structures were determined the synthetases were seen to be partitioned into two classes of 10 enzymes each, based on similarities in their catalytic cores.

Enzymes belonging to Class I function as monomers or homodimers (α_2). Their active sites contain a Rossmann nucleotide binding fold, also observed in dehydrogenases and other nucleotide-binding proteins. This fold consists of an alternating pattern of β-strands and α-helices. The nucleotide binding fold is split into two halves, with a variable-length polypeptide insertion known as CP1 between the halves. As described earlier, this polypeptide makes up the editing domain present in some enzymes. Comparison of Class I enzymes identified a signature sequence within the active site; this stretch of 11 amino acids is located in the first half of the nucleotide binding fold and ends in HIGH (the one-letter code for His–Ile–Gly–His). The second half of the fold contains another conserved sequence, KMSKS, which also makes critical contributions to the enzyme active site.

Class II AARSs were first distinguished by their lack of the Class I signature sequence, but this subfamily also has distinct features of its own. They are mostly α_2 dimers, although some are $\alpha_2\beta_2$ tetramers. The enzymes belonging to Class II have active sites composed of a seven-stranded antiparallel β-sheet with three α-helices. The characteristic Class II sequence motifs 1, 2, and 3 exhibit only minimal sequence conservation, but contribute a helix-loop-strand, strand-loop-strand, and strand-helix to the active site, respectively.

A mechanistic difference between the enzymes of Classes I and II is the site of initial amino acid attachment to the tRNA. Class I enzymes aminoacylate the 2′-hydroxyl of the terminal adenosine's ribose and most Class II enzymes use the 3′-hydroxyl. The amino acid subsequently migrates between the two positions. This early functional observation was rationalized once high-resolution enzyme–tRNA cocrystal structures were determined. Class I enzymes approach the tRNA acceptor stem from the minor-groove side of the RNA helix and are nearer the ribose 2′-hydroxyl, while Class II enzymes approach from the major-groove side of the tRNA, near the 3′-hydroxyl.

F. Mechanistic Clues from Structural Studies

Enzyme–tRNA cocrystal structures have also provided clues to the catalytic mechanisms used by AARSs. In the case of the Class I glutaminyl-tRNA synthetase (GlnRS) complexed with its cognate $tRNA^{Gln}$, the structure of the complex identified variations in the tRNA architecture compared with uncomplexed tRNA. GlnRS makes essential contacts with nucleotides in the $tRNA^{Gln}$ anticodon, and in the cocrystal structure these nucleotides were indeed splayed out from their normal positions to interact closely with amino acids of the enzyme. Furthermore, the acceptor end of the tRNA showed significant distortion compared to the expected helical structure. The first base pair of the acceptor helix was broken, and the 3′-end of the tRNA bent back in a hairpin conformation toward the rest of the helix (Fig. 3). Such a folded-back orientation of the acceptor end of $tRNA^{Gln}$ is necessary for the 3′-terminus to reach the glutaminyl adenylate in the enzyme active site. Furthermore, amino acid residues in the enzyme's active

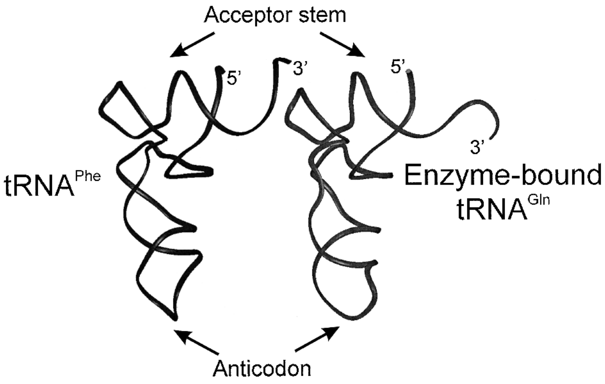


FIGURE 3 Distortion of $tRNA^{Gln}$ in complex with GlnRS. The cocrystal structure of GlnRS:$tRNA^{Gln}$ revealed a distortion at the 3′-end of the tRNA (right) in contrast to uncomplexed $tRNA^{Phe}$ (left). This uncoupling of the first acceptor helix base pair and folding back of the CCA end of the tRNA is apparently necessary for the tRNA to reach the glutaminyl adenylate bound in the enzyme active site.

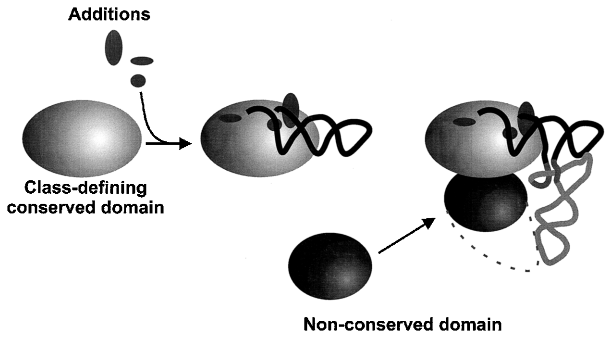


FIGURE 4 Domain organization of synthetases and tRNAs. The class-defining domain of the aminoacyl–tRNA synthetase contains conserved structural features and contacts the acceptor arm of tRNA. These contacts are mediated by additions to the class-defining catalytic domain. Appended nonconserved protein domains interact with other portions of the tRNA, including in many cases the anticodon (as indicated by the dotted line). [From Schimmel, P., and Ribas de Pouplana, L. (1995). *Cell* **81,** 983–986.]

site are inserted into the acceptor stem nucleotides to break the first base pair and facilitate the distortion of the end of the helix. Binding of the cognate anticodon may trigger this distortion, which is necessary for efficient transfer of the activated amino acid to the 2′-OH of the bound tRNA. Because the anticodon binding site is located a significant distance from the enzyme active site, tRNA recognition in this case has elements of a signal transduction mechanism.

G. Origins of Aminoacylation

As discussed earlier, AARSs have core catalytic domains that perform the functions of aminoacyl adenylate formation and transfer of the amino acid to the cognate tRNA. The sequences and structures of these domains also differentiate the enzymes as belonging to Class I or II. In addition to this class-defining active site domain, most AARSs also have one or more appended domains that are unique. These idiosyncratic domains often make specific contacts with recognition elements outside the tRNA acceptor stem, for example, at the anticodon or variable loop of the tRNA molecule (Fig. 4). In addition to the two-domain (or more) organization of the AARS enzymes, tRNAs can also be viewed as modular structures. As mentioned earlier, the acceptor stem and TΨC arm coaxially stack to form one portion of the L-shaped tRNA structure, while the D and anticodon arms stack to make the other tRNA arm (Fig. 2). The acceptor arm makes contacts with the catalytic core of the enzyme and contains the amino acid attachment site, while the anticodon, located on the second arm of the tRNA, is recognized by an appended domain.

Given this apparent segregation of functions between both AARS enzyme and tRNA substrate, it is interesting to consider whether an early protein synthesizing system might have used minimalist versions of both tRNA and AARS. Such a minimalist system is suggested by the observation that small RNA substrates based on the acceptor stem sequence but lacking the anticodon (such as the minihelices discussed earlier) are substrates for many AARSs. In the case of the tRNA substrate, a minihelix containing the amino acid attachment site would have been the earlier part of the tRNA, while the anticodon-containing arm could have emerged later as the template reading head. On the protein side of the aminoacylation reaction, the class-defining catalytic domain is envisioned as the ancestral enzyme, with domains making tRNA-specific contacts subsequently added.

H. Novel Functions of AARSs

In addition to their critical role in protein synthesis, it has become clear that AARSs are involved in several other cellular pathways. Some AARSs regulate their own transcription and translation, while others contribute to splicing activities in mitochondria. Nuclear aminoacylation of tRNAs by imported AARSs is thought to be a quality control mechanism to ensure that only mature, fully active tRNAs are released efficiently to the cytoplasm for protein synthesis. Programmed cell death (apoptosis) also appears to have an AARS component—human tyrosyl–tRNA synthetase can be proteolytically cleaved into two polypeptides with distinct cytokine activities, despite the lack of such activity in the full-length TyrRS. It is likely that in time many more nontranslational functions of AARS will be identified.

IV. AN OVERVIEW OF TRANSLATION

Translation of an mRNA message into its polypeptide product on the ribosome is a polymerization reaction, and can be divided into three phases: initiation, elongation, and termination. Initiation requires the assembly of the translational machinery from its individual components to form a complex that is primed for peptide bond formation. Formation of this initiation complex is usually the rate-limiting step in protein synthesis. Elongation is the sequential joining of amino acids by peptide bonds as dictated by the codons of the mRNA. Translation proceeds along the mRNA in the 5′ to 3′ direction, and the resulting polypeptide chain is synthesized in the amino-terminal to carboxy-terminal direction (N- to C-terminal). Termination occurs when a stop codon is reached in the message and the polypeptide is released from the ribosome, which is then recycled for translation of another protein. Each step has associated protein factors that facilitate substrate transport and ribosomal function.

The ribosome is a ribonucleoprotein particle, that is, it contains both RNA and protein components. The three RNA molecules (in prokaryotes, four in eukaryotes) and approximately 50 proteins that make up a ribosome are divided into two unequal subunits. Ribosomes and their subunits are characterized by their rate of movement (sedimentation) in intense centrifugal fields that greatly increase the force of gravity. Their sedimentation coefficients (S) are determined by size and shape. Bacterial ribosomes are 70S particles made up of 30S and 50S subunits. The smaller subunit (30S) contains the decoding site, where the mRNA–tRNA base-pairing interaction occurs. Peptide bond formation takes place on the larger ribosomal subunit (50S), at the peptidyl transferase center. When assembled, the subunits together form three binding sites for tRNA molecules. The A-site preferentially binds aminoacyl–tRNA, and is therefore where each incoming aminoacyl–tRNA complex is delivered for decoding. The nascent polypeptide chain is attached to the tRNA molecule in the P-site (for peptidyl–tRNA). Finally the E-site is the location of a deacylated tRNA molecule prior to its exit from the ribosome (E is for exit).

V. TRANSLATION INITIATION

Initiation of protein synthesis requires assembly of the ribosomal subunits, messenger RNA, and initiator tRNA at the start codon. This organization of translational components is facilitated by protein initiation factors (Fig. 5).

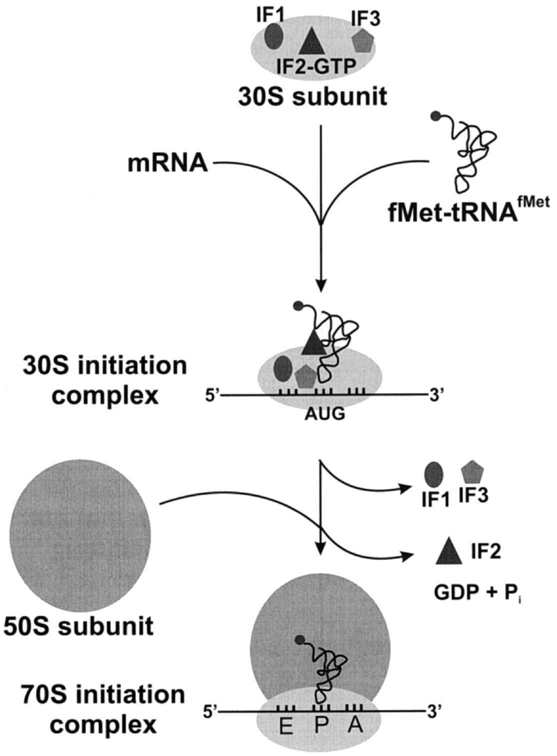


FIGURE 5 Translation initiation. In prokaryotes, three initiation factors are responsible for assembling the initiation complex prior to decoding of a message. The mRNA and initiator tRNA bind to the small ribosomal subunit in random order, with IF2 selectively binding initiator tRNA. Hydrolysis of IF2-bound GTP promotes formation of the 30S initiation complex. Initiation Factors 1 and 3 leave the complex, and the large ribosomal subunit binds to form the 70S initiation complex with release of IF2.

A. The Message

A combination of nucleotide signals identifies the beginning of an mRNA sequence to be translated into its protein product. The nucleotide triple AUG is the start codon that directs the ribosome to begin reading an mRNA and orients the message in the right frame (for example, . . . *CUA* GUG *CAC* C. . . rather than . . . *C UA*G UGC *ACC*. . . , which would be a different protein). However, AUG is also the codon for insertion of the amino acid methionine into the body of the polypeptide chain. What distinguishes the start AUG from other identical codons elsewhere in the message? A stretch of 3–10 nucleotides located about 10 nucleotides upstream (in the 5′-direction) from the start codon is called the Shine–Dalgarno sequence, after the researchers who identified it. This sequence is rich in A and G nucleotides, and is partially complementary to a short region of U and C nucleotides near the 3′-end of an RNA molecule embedded within the ribosomal small subunit. Such complementarity positions the incoming message properly on the ribosome, so that the start codon

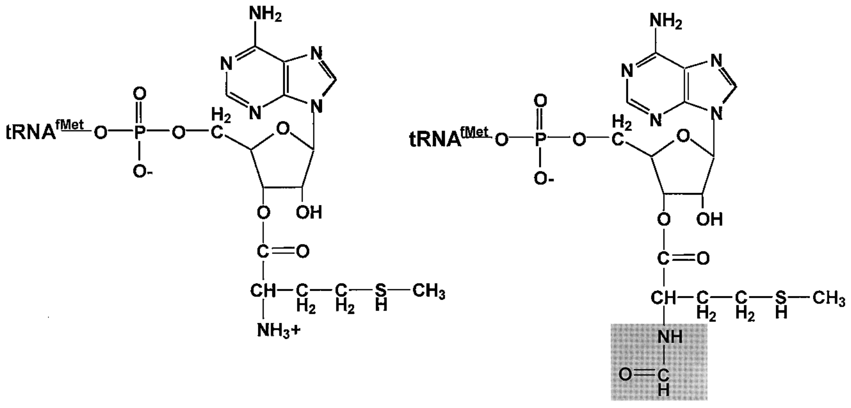


FIGURE 6 *N*-formyl methionyl–tRNA^{fMet}. Sequence and structural features target methionylated tRNA^{fMet} for formylation at the free amino group. This modification of the aminoacylated tRNA blocks the amino group and introduces an amide bond (highlighted in gray). Together with the uniquely rigid anticodon stem, these elements allow recognition by IF2 and placement of the initiator molecule in the ribosomal P-site for interaction with the AUG start codon of mRNA.

is in the ribosomal decoding site for initiation of protein synthesis. Once translation begins, the rest of the codons in the message need to be read, so the base-pairing interaction between mRNA and rRNA at the Shine–Dalgarno sequence must be transient.

B. Initiator tRNA

The tRNA that recognizes the AUG start codon and the amino acid attached to it are also distinct from the Met–tRNA^{Met} base-paired to internal AUG codons. This initiator tRNA contains the CAU anticodon necessary for recognition of the start codon, but several sequence and structural differences distinguish it from the elongator tRNA that inserts methionine into internal positions of the polypeptide. In bacteria, both elongator and initiator tRNAs are aminoacylated by methionyl–tRNA synthetase, but the methionylated initiator tRNA undergoes further processing prior to its transport to the ribosome. The enzyme fMet–tRNA transformylase modifies the amino group of the tRNA-attached methionine residue, using N^{10}-formyltetrahydrofolate as a formyl donor. One feature that the transformylase enzyme recognizes is a mismatched base pair at the first position of the initiator tRNA (tRNA^{fMet}) acceptor stem. The elongator tRNA (tRNA^{Met}) molecule has a canonical G:C base pair at the first position of the acceptor stem, while tRNA^{fMet} contains a C:A pair. Base substitutions that produce a strong base pair at this position in tRNA^{fMet} significantly decrease formylation of Met–tRNA^{fMet}.

Modification of the aminoacylated initiator tRNA to produce formyl–methionine–tRNA^{fMet} blocks the amino group of methionine and introduces an amide bond (Fig. 6). The lack of a free amino group in fMet–tRNA^{fMet} prevents its insertion into a protein anywhere but at the N-terminal position. Furthermore, the presence of an amide bond targets fMet–tRNA^{fMet} for the P-site of the ribosome. Only the initiator tRNA enters the P-site directly—all other aminoacyl–tRNAs enter the ribosome at the A-site and are moved to the P-site after peptide bond formation. This targeting is achieved both by the presence of the amide bond of fMet–tRNA^{fMet} and by the uniquely rigid anticodon stem of the initiator tRNA, which contains three G:C base pairs. Progressive substitution of these three G:C pairs has been shown to weaken binding of the initiator tRNA to the P-site.

C. Initiation Factors

Assembly of the ribosomal subunits, mRNA, and initiator tRNA into a complex ready for protein synthesis requires several proteins called initiation factors. In prokaryotes, three initiation factors (IFs) transiently associate with the components of the translational machinery: IF1, IF2, and IF3. (In eukaryotes, more factors are required but the overall initiation process is similar with a few exceptions described below.) Table II summarizes the properties of *E. coli* initiation factors as well as protein factors involved in elongation and termination.

Biochemical studies determined that each initiation factor has a distinct high-affinity binding site on the 30S subunit. Based on *in vivo* concentrations and affinities, the IFs most probably do not exist free in solution but are predominantly bound to 30S subunits. When cellular concentrations of mRNA and fMet–tRNA^{fMet} are high enough, these RNAs bind the small subunit (containing

TABLE II Properties of *E.coli* Translational Factors

	Protein factor	Size in *E. coli* (kD)	Function
Initiation	IF1	9	Stimulates IF2/IF3 binding and functions
	IF2	97	GTPase, binds fMet–tRNA^{fMet}
	IF3	22	Prevents subunit association
Elongation	EF-Tu	43	GTPase, transports AA–tRNA to ribosome
	EF-G	74	GTPase, stimulates translocation of tRNA on ribosome
	EF-Ts	77	Nucleotide exchange factor (EF-Tu: GDP to EF-Tu:GTP)
Termination	RF1	36	Promotes polypeptide release at stop codons UAG and UAA
	RF2	38	Promotes polypeptide release at stop codons UGA and UAA
	RF3	46	GTPase, stimulates RF1 and RF2 function
	RRF	20	Promotes dissociation of post-termination complex

a single molecule of each IF) in random order. In this ternary complex, the mRNA and tRNA are not in contact with each other, but a kinetic rearrangement triggered by the IFs positions the initiator tRNA in the part of the P-site contributed by the 30S particle and promotes the first codon–anticodon interaction. This 30S initiation complex can either dissociate into its individual components or bind the 50S subunit (with IF2-mediated hydrolysis of GTP) to form a 70S initiation complex.

Each factor plays a particular role in translation initiation, acting together to form 30S initiation complexes that can proceed toward elongation. The precise function of IF1 is not known, although it does stimulate the binding and activities of IF2 and IF3. It binds to 30S subunits directly, and in combination with the other factors promotes the formation of 30S initiation complexes. Once the 30S complex is formed, IF1 is ejected along with IF3. To date there is no conclusive evidence that the IFs affect the Shine–Dalgarno interaction between the mRNA and rRNA.

The most active role in initiation seems to belong to IF2, which has binding sites for fMet–tRNA^{fMet}, GTP, and both ribosomal subunits. Initiation Factor 2 promotes the association of fMet–tRNA^{fMet} with the small subunit and, in particular, recognizes the blocked amino group of this tRNA. The activities of IF2 are partitioned between two domains of the protein—the C-terminal domain is thought to be responsible for initiator tRNA binding, while the central domain contains a GTPase function. Which portion of IF2 binds the 30S subunit remains to be determined. Hydrolysis of GTP to GDP occurs only upon 50S binding to the ternary complex; IF2 has no GTPase activity in the absence of the ribosome. As GTP hydrolysis is the final step in initiation, a conformational change in the ribosome is thought to eliminate IF2 from the initiation complex, further orient the initiator tRNA in the P-site, or otherwise contribute to a kinetic proofreading mechanism.

Initiation Factors 2 and 3 have seemingly opposing functions. While IF2 promotes the binding of tRNA to the 30S subunit, IF3 can be considered the subunit "antiassociation" factor because it increases the rates of subunit exchange and complex dissociation. In fact the two functions cooperate to facilitate formation of the correct 30S initiation complex—IF2 preferentially enhances the binding of the amino-blocked initiator tRNA and IF3 specifically increases the rate of noninitiator tRNA dissociation from the ternary complex. Initiation Factor 3 also contributes to the fidelity of translation by confirming the codon–anticodon interaction on the 30S subunit.

D. Eukaryotic Initiation

Many more protein factors are involved in eukaryotic initiation; some systems contain more than 10 initiation factors. Particular features of translation initiation are also different in the higher organisms. Most notably, prokaryotic ribosomes can initiate internally on an mRNA (even on circular RNAs), while in eukaryotes a "preinitiation" complex binds to the 5′-end of the mRNA and then progresses to an initiation complex. Eukaryotic mRNAs are capped at their 5′-end with a 7-methylguanosine triphosphate structure, and one of the eukaryotic initiation factors binds this capped end. The preinitiation complex then moves along the mRNA and initiates translation at the first AUG codon it comes to. Consistent with this scanning mechanism is the observation that eukaryotic mRNAs do not contain Shine–Dalgarno-like sequences.

VI. ELONGATION

The heart of protein biosynthesis lies in the elongation cycle, with its sequential decoding of mRNA codons to assemble the useful portion of the polypeptide. Elongation can be further broken down into three phases—aminoacyl–tRNA decoding, peptide bond formation, and translocation of the new peptidyl–tRNA (Fig. 7).

A. Decoding According to Base Pairing

Comparison of the anticodon of the incoming AA–tRNA with the corresponding mRNA codon takes place at the

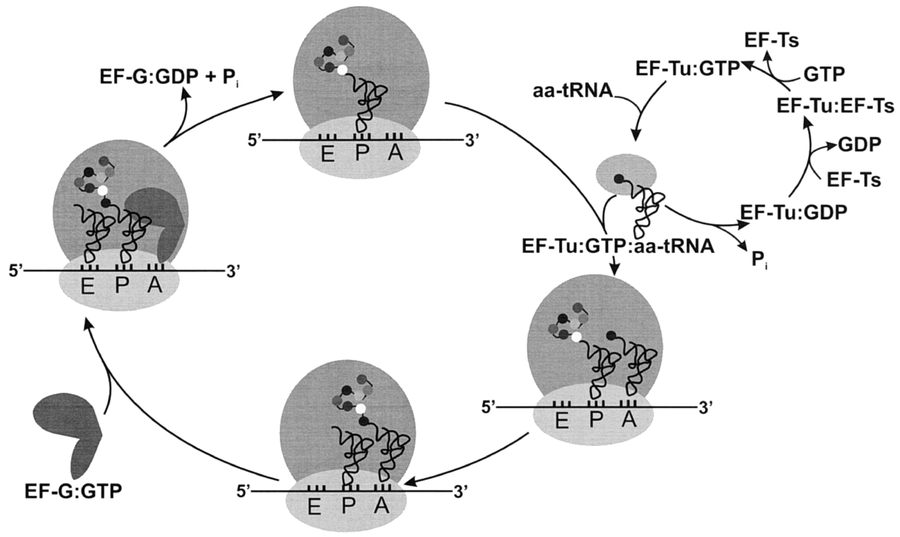


FIGURE 7 The elongation cycle. Aminoacylated tRNAs are transported to the ribosome by EF-Tu, and they are positioned in the ribosomal A-site upon hydrolysis of EF-Tu-bound GTP. Nucleotide exchange is catalyzed by EF-Ts. Following peptide bond formation, the GTPase EF-G triggers translocation of the peptidyl–tRNA from the A-site to the P-site; the empty (deacylated) tRNA exits the ribosome by way of the E-site. The mRNA moves the length of one codon in the 3′ direction, probably pulled through the ribosome by tRNA translocation.

decoding center of the ribosome, located on the small subunit. Peptide bond formation occurs on the large ribosomal subunit, at the peptidyl transferase center. Thus, this segregation of functions parallels the two arms of the tRNA. The anticodon portion of the tRNA binds to the small subunit, where the genetic message is read. The acceptor arm of tRNA (with its attached amino acid) contacts the large subunit, where catalysis occurs.

Although the synthesis of a peptide bond is the key step in translation, this is the easiest part of protein synthesis. Once the amino group of an aminoacyl–tRNA is properly positioned close enough to the carbonyl group of a peptidyl–tRNA, peptide bond formation through nucleophilic attack is energetically favorable. The ribosome can be considered as a single enzyme whose function is to catalyze peptide bond formation.

B. Elongation Factors

Addition of each incoming amino acid requires the cooperation of three elongation factors. Elongation factor Tu (EF-Tu) is the most abundant protein in *E. coli*, with about 100,000 copies per cell, or 5% of the cell's protein. This protein is a GTPase, and the EF-Tu:GTP complex specifically binds aminoacyl–tRNAs (AA–tRNAs). Formation of the ternary complex (EF-Tu:GTP:AA–tRNA) protects the ester bond (linking the amino acid to its cognate tRNA) from hydrolysis, and transports the AA–tRNA to the ribosomal A-site. Once the correct codon–anticodon interaction is confirmed, ribosome-triggered hydrolysis of EF-Tu-bound GTP occurs. EF-Tu:GDP is then released from the ribosome, and the AA–tRNA occupies the A-site. While EF-Tu transports all elongator tRNAs aminoacylated with natural amino acids to the ribosome, this factor has negligible affinity for formylated or nonformylated $tRNA^{fMet}$. The unpaired first position in the $tRNA^{fMet}$ acceptor stem helix apparently is a negative recognition element for EF-Tu:GTP, because this element prevents the initiator tRNA from pairing with internal AUG or GUG codons.

The elongation factor EF-Ts is a nucleotide exchange factor that regenerates active EF-Tu:GTP (from EF-Tu:GDP) for binding subsequent AA–tRNAs following GTP hydrolysis. Before their functions were known, elongation factors Tu and Ts were named for their observed thermal stabilities *in vitro*—Tu indicates that this protein is <u>T</u>emperature <u>u</u>nstable, while Ts stands for <u>T</u>emperature <u>s</u>table. In eukaryotes, the two subunits of elongation factor EF-1 perform the functions of EF-Tu and EF-Ts.

Once the A-site is occupied by the incoming AA–tRNA, nucleophilic attack on the peptidyl–tRNA by the AA–tRNA occurs. The condensation reaction produces a new peptide bond and lengthens the polypeptide chain by one amino acid (Fig. 8). As a result, the growing protein is now attached to the A-site tRNA; this addition to and transfer of the polypeptide chain is called transpeptidation.

Following formation of the peptide bond, a major rearrangement of components in the functional center of the ribosome must take place. Because the most recently entered tRNA has become the peptidyl–tRNA, it must

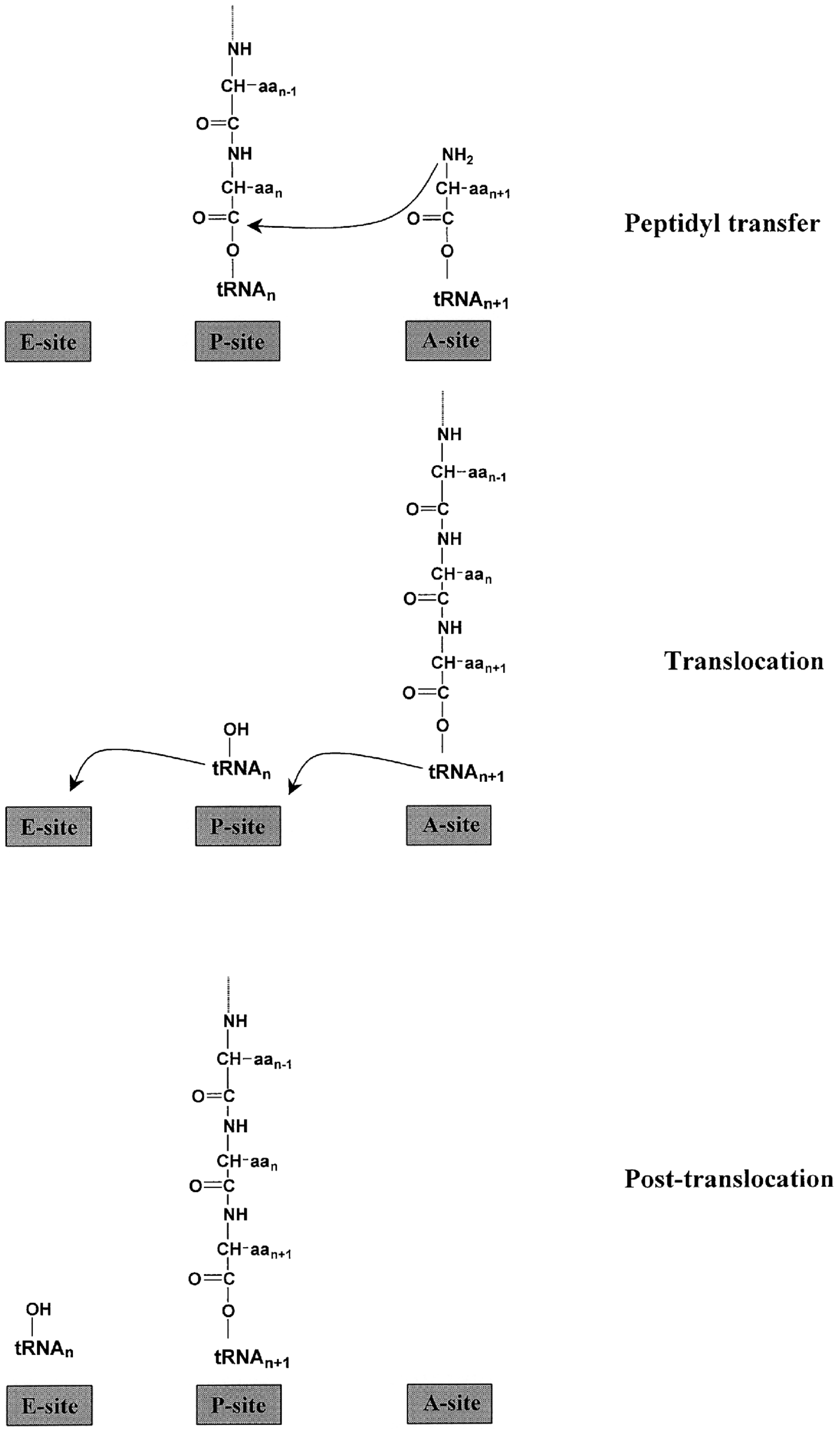


FIGURE 8 Steps in elongation. With the peptidyl–tRNA bound in the P-site and the incoming aminoacyl–tRNA in the A-site, the peptidyl transferase activity of the large ribosomal subunit catalyzes peptide bond formation. The growing polypeptide chain is then attached to the A-site-tRNA, and the deacylated tRNA is in the P-site. Elongation factor G facilitates translocation of the peptidyl–tRNA to the P-site and the empty tRNA to the E-site prior to its release from the ribosome.

be moved from the A-site to the P-site. The former peptidyl–tRNA has been deacylated and needs to vacate the P-site. Finally the mRNA must move three nucleotides further in the 3′-direction so that the next codon can be read. The concerted movement of tRNAs and mRNA at the end of each elongation round is called translocation, and is catalyzed by elongation factor G (EF-G), another of the GTPase proteins in the translational machinery.

VII. TERMINATION

When the mRNA stop codon is reached, the fully synthesized protein does not simply fall off the ribosome. Release factors (RFs) are the protein assistants that recognize the presence of a stop codon in the ribosomal A-site and trigger cleavage of the polypeptide from the P-site tRNA (Fig. 9). In prokaryotes, RF1 hydrolyzes the protein at stop codons UAG and UAA, while RF2 recognizes stop codons UGA and UAA. Release factor RF3, a GTPase, stimulates the activities of the codon-specific RFs and also promotes the dissociation of RF1 or RF2 from the ribosome once the newly synthesized protein is released. In eukaryotes, only two RFs appear to trigger polypeptide release—one that combines the functions of RF1 and RF2 (recognizing all stop codons), and a GTPase like RF3.

The observation that prokaryotic RF1 and RF2 exhibit specificity for different RNA sequences suggests that direct contacts are made between the release factors and the mRNA. These protein–RNA interactions contrast with the codon–anticodon contacts defined by Watson–Crick base-pairing rules, and hint at a mechanism in which the proteins mimic bound tRNA.

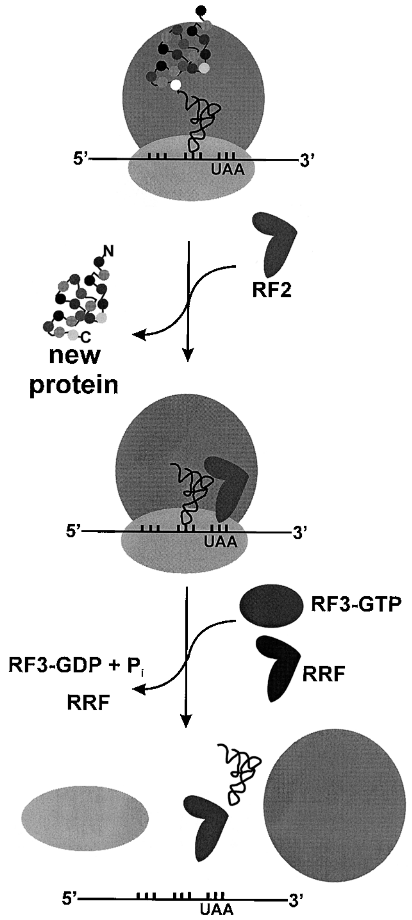


FIGURE 9 Termination of protein synthesis and ribosome recycling. In prokaryotes, RF1 hydrolyzes the newly synthesized protein at stop codons UAG and UAA, while RF2 recognizes stop codons UGA and UAA. The GTPase RF3 stimulates release of either RF1 or RF2. In eukaryotes a single protein recognizes all stop codons. The final step of translation is dissociation of the inactive 70S complex, stimulated by the ribosome recycling factor (RRF).

VIII. RIBOSOME RECYCLING

Following release of the synthesized protein, the ribosome contains an empty tRNA in the P-site or E-site, and mRNA is still bound with the stop codon in the A-site. This arrangement of components is the post-termination complex. A protein known as the ribosome recycling factor (RRF) promotes the dissociation of this complex in preparation for translation of another gene. This recycling of translation components has been called the fourth step of protein synthesis, after formation of the initiation complex, elongation, and release of the full-length polypeptide at the stop codon. If the post-termination complex is not dissociated, the ribosome remains bound to the mRNA and unprogrammed translation can be reinitiated downstream of the stop codon. In fact, RRF is necessary for prokaryotic cell growth. Dissociation of the post-termination complex by RRF is likely facilitated by EF-G and IF3, and this completion of the protein synthesis cycle frees the translational machinery to read another mRNA.

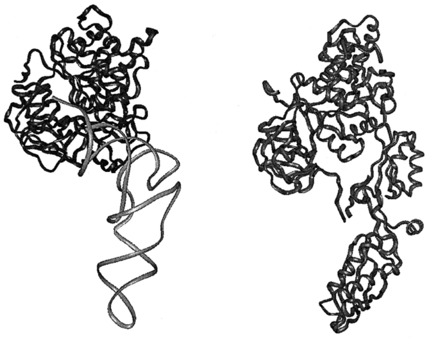
IX. MOLECULAR MIMICRY BY TRANSLATIONAL FACTORS

The mechanisms by which several of the translational factors act in protein synthesis have been suggested based on detailed structural analyses. Elongation factor Tu was the first factor for which an X-ray crystallographic structure was determined, and was also the first GTP-binding protein whose structure was elucidated. This protein is organized in three structural domains. Domain 1 is the GTP-binding domain (G-domain) consisting of a nucleotide

binding fold now known to be typical of G-proteins. Domains 2 and 3 are β-barrels, held together by strong interdomain contacts such that the two domains act as a single structural unit. Crystal structures of the many functional forms of EF-Tu have been determined—EF-Tu:GDPNP (a nonhydrolyzable analog of GTP), EF-Tu:GDP, and complexes with EF-Ts or AA–tRNA. The orientation of Domain 1 with respect to Domain 2/3 varies greatly among the functional forms of EF-Tu. The two parts of the protein rotate by 90° when GDP–GTP exchange is catalyzed by EF-Ts; this conformational change is due to two switch regions in the G-domain. By altering small secondary structure elements, these regions trigger long-range effects.

The crystal structure of the ternary complex with Phe–tRNA (EF-Tu:GDPNP:Phe–tRNA^{Phe}) demonstrated that the EF-Tu structure in the ternary complex is similar to that in the EF-Tu:GDPNP structure. Thus, binding of AA–tRNA does not alter the EF-Tu conformation. The ternary complex is quite elongated (Fig. 10), with the tRNA anticodon pointing away from EF-Tu, and close contacts are observed only between the factor and the T-stem, 3′-CCA–AA, and 5′-phosphate of the AA–tRNA. The Phe–tRNA^{Phe} structure is also not significantly altered upon binding to EF-Tu.

When the crystal structure of EF-G:GDP was solved, it revealed a surprising and elegant structural feature. Elongation factor G consists of five structural domains, and from sequence comparisons Domains 1 and 2 were expected to be similar in conformation to EF-Tu Domains 1 and 2. This conformational "mimic" does indeed occur. Domains 3 and 5 of EF-G contain protein folds similar to some ribosomal proteins whose structures are known, while Domain 4 adopts an unusual fold. This domain is elongated and points away from the rest of the protein.



EF-Tu:GDPNP:tRNA^{Cys} **EF-G**

FIGURE 10 Comparison of elongation factor structures. The crystal structures of the EF-Tu:GDPNP:tRNA^{Cys} ternary complex (left) and EF-G (right) revealed that Domains 3, 4, and 5 of EF-G mimic the conformation of EF-Tu-bound tRNA. Several other translational factors have been determined or predicted to similarly mimic the tRNA structure.

Upon comparison with the EF-Tu ternary complex structure, EF-G was seen to have a conformation similar to that of the ternary complex, with Domains 3, 4, and 5 of EF-G mimicking AA-tRNA.

The remarkable mimicry of tRNA by portions of EF-G suggested a mechanism by which the factor might be facilitating translocation. It is attractive to imagine that EF-G actively "chases" the peptidyl–tRNA into the ribosomal P-site because it mimics A-site-bound tRNA. Furthermore, the sequence of events dictates that, immediately after translocation, EF-G:GDP is released and another ternary complex (EF-Tu:GTP:AA–tRNA) enters the ribosomal A-site. Therefore the departure of EF-G:GDP leaves behind a binding site preformed for acceptance of the ternary complex.

Perhaps it should not be surprising that such structural mimicry is apparently used more than once in protein synthesis. Recent structure studies have determined that eRF1 (eukaryotic release Factor 1, which recognizes all stop codons) and RRF also contain domains that closely resemble tRNA. Although there are now several examples of factors that mimic the shape of tRNA, the factors themselves are structurally distinct from one another. This suggests a convergent appropriation of the tRNA structure, rather than a gene duplication event. In contrast, sequence comparisons of the G-protein factors EF-Tu, EF-G, IF2, and RF3 showed that the structural folds seen in Domains 1 and 2 of EF-Tu and EF-G are present in IF2 and RF3 and suggest that they will be present in the other factors as well.

Not only do several of the translational factors mimic the overall shape of tRNA, but also, even at the atomic level, a protein can copy the recognition events of the anticodon–codon interaction. The prokaryotic release factors RF1 and RF2 terminate polypeptide release at stop codons UAG/UAA and UGA/UAA, respectively. The basis for codon discrimination was localized to a tripeptide motif within the two release factors, with the first and third amino acids specifying the second and third base positions of the codon. For example, the first position (proline) of the "peptide anticodon" dictates that only A is recognized while serine is permissive to either A or G. Such relaxed recognition may be similar to wobble pairing in codon–anticodon interactions. This model for stop codon recognition was tested by generating protein variants that switched the codon specificity between release factors, recognized all stop codons, or were restricted to the UAA termination signal. The nucleotide feature being used for discrimination is likely the C2 amino group of G (where there is only a proton in A). The release factors are therefore making direct contact with the mRNA stop codons in a way that mimics the base-pairing of RNA–RNA interactions.

X. TRANSLATIONAL ACCURACY

A. Types of Potential Errors

Given the complexity of protein synthesis, there are many steps at which mistakes can occur, and it is remarkable that the error rate is as low as it is. The two principal categories of translational mistakes are missense errors, in which one amino acid is substituted for another or a stop codon is not recognized, and errors of processivity, in which frameshifting or premature termination occur.

Missense errors can take place because of misacylation of a tRNA by an AARS, resulting in insertion of an incorrect amino acid despite a correct codon–anticodon match. This potential problem is counteracted by active editing by AARSs (see earlier), which is critical for maintaining translational accuracy. Errors can also occur at the level of the codon–anticodon interaction, such that a noncognate AA–tRNA is selected at the ribosome, again resulting in amino acid misincorporation. The relative abundance of cognate vs. noncognate AA–tRNA affects the accuracy of amino acid incorporation. In particular when the next amino acid specified by the mRNA is unavailable, the frequency of missense errors increases. A special case of incorrect AA–tRNA selection is that of stop codon readthrough, in which an amino acid is inserted where the protein should actually be terminated. This type of error results in a carboxy-terminal extension (added to the nascent protein) that continues until the ribosome reaches the next stop signal.

Frameshift mutations occur because of a nucleotide insertion or deletion in a protein's gene. Errors in reading frame also occur when the ribosome "slips" along the mRNA in such a way that the sequence is not read in triplets corresponding to the codons of the message, but the ribosome moves two or four nucleotides instead. These "slips" are -1 or $+1$ frameshifts, respectively. Fortunately the ribosome usually encounters a stop codon shortly after the frameshift; this minimizes the effect of the mutation. (Although frameshifting can be the unintended result of translational inaccuracy, "programmed" frameshifts also take place in special situations. These produce alternate polypeptides from a single mRNA.) Premature termination of a protein can occur when a nucleotide substitution produces a stop codon in the middle of the gene sequence. The peptidyl–tRNA may also dissociate prematurely from the ribosome before the stop codon is reached.

B. Suppressor tRNAs

For several of these error scenarios, faithful translation of the genetic message is accomplished through the action of suppressor tRNAs. These are the result of advantageous mutations in tRNA genes, typically in the anticodon, that allow them to recognize a misplaced stop codon or unintended frameshift. Suppressor tRNAs may be aminoacylated according to the amino acid specificity of their parent sequence, so therefore they insert this amino acid into the polypeptide at the location of the mRNA aberration. Alternatively, they may be misacylated because of the change in the anticodon sequence. As long as the inserted amino acid is not detrimental to the synthesized protein, a suppressor tRNA rescues the gene mutation. For example the *E. coli su6* suppressor tRNA inserts leucine into the growing polypeptide chain at the position of a premature UAA stop codon.

Missense suppressors simply substitute one amino acid for another to correct a mutation in the mRNA. Transfer RNAs that suppress $+1$ frameshifts typically contain an extra base in the anticodon and read a four-nucleotide codon, thereby restoring the correct translational frame.

Suppressor tRNAs are of particular interest as more examples of the genetic basis for specific diseases are found. In many cases, missense or nonsense mutations in genes for essential proteins correlate with diseases such as cancer, cystic fibrosis, and amyotropic lateral sclerosis (ALS). In fact, mutations in the tumor suppressor gene *p53* are the single greatest cause of human cancers. One particular *p53* hotspot is at codon 213, where nonsense mutations have been identified in human colorectal, gastric, ovarian, and breast cancers. Suppressor tRNAs designed to selectively rescue these mutations by reading through the premature termination codon could be effective therapeutic agents in a gene therapy approach. Several challenges remain, however—the suppressor tRNA must be introduced into the affected cells, must be transcribed at high enough levels to be effective, must also be aminoacylated at high levels, and must not excessively read through stop codons that specify the normal termination of translation for other genes. These are active areas of research in several laboratories.

C. Selenocysteine Insertion at UGA Codons

The trace element selenium is present in a number of proteins in the form of cotranslationally inserted selenocysteine (Sec). This insertion occurs when a unique selenocysteinyl–tRNA^{Sec} recognizes an internal UGA codon. Selenocysteine has therefore been called the "21st amino acid" and its insertion at a nucleotide triplet that is normally a termination signal represents an expansion of the genetic code.

Insertion of selenocysteine requires several adaptations of the translational machinery. First, tRNA^{Sec} is aminoacylated with serine by seryl–tRNA synthetase. The serine attached to this misacylated species is then converted

to selenocysteine by the enzyme selenocysteine synthase, which uses selenophosphate as a donor. Elongation factor Tu does not bind Sec–tRNA^{Sec} as it does other elongator AA–tRNAs; instead, a unique protein SELB transports Sec-tRNA^{Sec} to the ribosomal A-site. SELB is specific for Sec–tRNA^{Sec}, rejecting other AA–tRNAs including Ser–tRNA^{Sec}. The final novel feature of selenocysteine insertion is the mechanism of mRNA recognition. The UGA triplet can be used in the same organism as either a selenocysteine or a stop codon. The sequence context determines its recognition as a selenocysteine codon. The ternary complex SELB:GTP:Sec–tRNA^{Sec} recognizes a stem-loop structure immediately 3′ (downstream) from a UGA codon that is read as selenocysteine. This structural feature of the mRNA is specifically bound by the carboxyl-terminal portion of SELB, while other regions of the protein are highly homologous to EF-Tu as expected. Insertion of selenocysteine into polypeptides therefore requires formation of a quaternary SELB:GTP:Sec–tRNA^{Sec}:mRNA complex. This is in contrast to all other elongation steps in protein synthesis, which proceed through ternary complexes.

D. Degradation of Incomplete Polypeptides

One consequence of the necessary accuracy in protein synthesis is the release of peptidyl–tRNA molecules representing incomplete translation products. These products can be the result of ribosome stalling, a premature stop codon that is not suppressed, or detection by the ribosome of noncognate tRNA present in the decoding center. It has been estimated that this "drop-off" might result in a truncated polypeptide chain approximately 10% as often as the full-length protein. Not only is this a waste of amino acids resulting in useless products but also tRNA molecules are sequestered and unavailable for translation of other genes.

The incomplete peptidyl–tRNAs are substrates in bacteria and yeast for the enzyme peptidyl–tRNA hydrolase, which cleaves the ester bond between the tRNA and its attached polypeptide. Peptidyl–tRNA hydrolase therefore removes the useless (and potentially harmful) protein fragments and recycles the tRNAs. Interestingly, although the initiator tRNA (fMet–tRNA^{fMet}) mimics a peptidyl–tRNA by virtue of its *N*-formyl group, peptidyl–tRNA hydrolase does not recognize the initiator as a substrate. The hydrolase bypasses fMet–tRNA^{fMet} because of the presence of structural features unique to fMet–tRNA^{fMet}.

E. A Dual-Function RNA

A particular challenge arises when an mRNA lacks in-frame stop codons due to deletion or degradation. Bacteria use a unique tRNA–mRNA hybrid (tmRNA, also called 10Sa RNA) to remove the resulting partially synthesized proteins and free the ribosomal subunits. The 5′- and 3′-ends of this molecule resemble alanyl–tRNA, while the central portion encodes a peptide "tag." Alanyl–tRNA synthetase aminoacylates the tmRNA, which is then transported to the stalled elongation complex by EF-Tu. After attaching alanine to the end of the growing nascent polypeptide, the ribosome switches from the truncated mRNA to the tmRNA in a mechanism called *trans*-translation. The ribosome adds 10 additional amino acids to the end of the protein according to the tmRNA sequence. The resulting tagged protein is released and degraded, as the tag is a recognition signal for several proteases.

F. mRNA Surveillance

Eukaryotes use a surveillance mechanism to identify mRNAs with mutations (typically premature termination codons) or processing errors (such as incorrect splicing). Once detected, the aberrant messages are degraded to prevent the synthesis of truncated proteins. Recent evidence suggests that these mRNAs must be at least partially translated to determine whether a stop codon is in its proper context. In mammalian systems, the translating ribosome is proposed to measure the distance between the final splicing junction and the termination signal—if they are within 50 nucleotides of one another, termination is allowed to proceed. If they are further apart, either because of a misplaced stop codon within the mRNA reading frame or a splicing error, the mRNA is targeted for rapid degradation. How the ribosome recognizes this distance is not yet known. This mRNA surveillance is also called nonsense-mediated decay because the majority of mutational errors result in premature termination codons.

G. Accuracy Mechanisms

Despite the types of translational errors described above, mRNA-directed protein synthesis is remarkably accurate. How is it that the ribosome and translational factors are able to achieve such faithful transmission of genetic information? One way to describe the specificity of cognate over noncognate AA–tRNAs is termed the "kinetic proofreading" mechanism. One can imagine that selection of an EF-Tu:GTP:AA–tRNA ternary complex by the ribosome during elongation can be considered a "scanning" step. Depending on the codon–anticodon interaction, the AA–tRNA will either bind irreversibly in the ribosomal A-site (in the case of the cognate AA–tRNA), or dissociate from the ribosome before or after GTP hydrolysis (noncognate AA–tRNA). In this model, the rate of EF-Tu-triggered GTP hydrolysis is the same for cognate and noncognate substrates. However, because the cognate

AA–tRNA spends more time in complex with the ribosome, GTP hydrolysis is likely to occur prior to dissociation of the tRNA, promoting tight binding of this tRNA in the A-site prior to peptide bond formation.

In addition to passive kinetic proofreading, selection of the correct codon–anticodon interaction is proposed to trigger a conformational change in the ribosome that is transmitted to EF-Tu. Hydrolysis of GTP is then accelerated for the cognate compared to the noncognate substrate. Recent structural evidence shows that structural rearrangements do occur upon binding of the cognate AA–tRNA, suggesting that selection of the correct substrate depends on an induced fit mechanism. How the codon–anticodon interaction is detected by the ribosome is not currently understood.

H. Ribosomal Contributions to Accuracy

Naturally occurring mutations have been identified in ribosomal proteins and rRNA that increase or decrease translational accuracy. These have provided clues as to which regions of the ribosome are involved in proofreading, although detailed mechanisms are not known. The small subunit, which contains the decoding site, has several proteins that are likely involved in controlling missense and processivity errors. For example, mutations have been identified in proteins S4 and S5 that reduce the level of translational fidelity (these are called ribosomal ambiguity mutations). In contrast, S12 mutants increase the accuracy of translation by conferring resistance to streptomycin, an error-causing antibiotic.

The small subunit 16S rRNA also contains regions that appear to be involved in translational accuracy. In particular, substitutions at nucleotides in the so-called 530 loop lead to changes in the error rate of protein synthesis. This rRNA loop is known to be spatially near the proteins S4, S5, and S12, and effects of nucleotide changes in the 530 loop parallel those observed for the proteins. Thus, the rRNA and proteins in this part of the small subunit together act as a proofreading domain. Substitutions at some of these nucleotides increase the rate of missense or frameshift errors, while others are detrimental because they prevent binding of the EF-Tu:GTP:AA–tRNA ternary complex. Still other mutations actually increase the accuracy of translation, in that they make the ribosomes resistant to error-inducing antibiotics.

The toxins α-sarcin and ricin inactivate ribosomes by altering a highly conserved sequence in 23S-like rRNAs (the sarcin–ricin loop). Elongation factors Tu and G bind near this loop, suggesting that it is involved in AA–tRNA selection and translocation. Protein L6 and the sarcin–ricin loop are involved in translational accuracy control, as suggested by the mutations in these components that increase fidelity, apparently by slowing down translation to allow more thorough proofreading.

XI. THE RIBOSOME

The bacterial ribosome has been the subject of intense study for several decades. Although the general mechanisms of protein synthesis (as outlined earlier) are reasonably well understood, only recently have structures emerged which make a molecular description of ribosome function appear possible. Because of the high degree of functional and sequence conservation between bacterial and eukaryotic components of the ribosome, structural information obtained using bacterial ribosomes is expected to contribute to a universal understanding of ribosomal architecture.

A. Composition of the Ribosome

The *E. coli* ribosome contains 3 ribosomal RNA (rRNA) molecules and about 50 proteins, divided between two unequal subunits. (The number of proteins is not entirely certain, as some proteins are loosely associated with the ribosome but may not be integral to its function.) The small subunit sediments as a 30S particle, which has a single 16S rRNA molecule and 21 proteins, S1–S21 (S in this case indicates a small subunit protein). The large subunit has two rRNAs (5S and 23S) and at least 34 proteins, with L-prefixes. In the cell, ribosomal subunits spontaneously assemble from their protein and rRNA components and are brought together as a functional complex with mRNA and tRNA at the initiation step of translation. Individual subunits and functional ribosomes can also be reconstituted *in vitro* under the right salt and temperature conditions. The determination of these conditions has greatly facilitated structural studies of the ribosome. For example, subunits can be reconstituted in the absence of one or more proteins and then assayed for structural integrity and function. Such experiments determined which proteins are essential, either because they are needed for subsequent binding of other proteins or because they are necessary for ribosomal function.

B. Sequence Conservation of Ribosomal Components

Nucleotide sequences are now known for rRNAs from hundreds of organisms. Comparisons of these sequences allowed predictions of the secondary structures of these RNAs (which regions are single stranded and which base pair to form helices). Interestingly, the rRNA secondary structures are evolutionarily conserved, while the individual nucleotides that make up these structures often are not.

Only small stretches of rRNA are identical in the majority of organisms, and these are therefore predicted to be critical for ribosomal function.

More sequence conservation is observed for ribosomal proteins than for rRNA. Structural similarities are significant enough that cross-species RNA interactions were shown possible for proteins L1, L11, and S15. At a minimum, then, the RNA-binding features of these proteins are conserved.

C. Structural Studies of the Ribosome

The ribosome's size challenged structural studies since its first observation in electron micrographs nearly 50 years ago. With a molecular mass of 2500 kD and dimensions about 250 Å on a side, the 70S ribosome is small for electron microscopy but immense for structural studies (such as NMR and X-ray crystallography) typically applied to single molecules.

The size and complexity of the ribosome have resulted in the development and application of many different experimental approaches to probe the arrangement of proteins and rRNA. Electron microscopy hinted at the overall shape of individual subunits and the assembled ribosome. The delineation of these shapes has changed little even with current high-resolution structures. The generation of antibodies against ribosomal proteins, rRNA termini, and modified bases allowed identification of their respective locations on the ribosome surface by means of immunoelectron microscopy. Specific sites of protein–rRNA interactions were determined by chemical modification experiments, which generate a "footprint" of protein binding according to the differential reactivity of rRNA nucleotides in the presence and absence of the protein. Similar modification experiments also confirmed the secondary structure predictions made for rRNAs, using enzymes and chemicals that react selectively with bases in either single-stranded or helical regions.

More detailed information was obtained about the orientation of proteins and rRNA, using reagents that generate covalent cross-links between ribosomal neighbors. Protein–protein, protein–RNA, and RNA–RNA cross-links provided valuable constraints for model-building studies by limiting the distance that these components could be placed from one another according to the length of the cross-linker. A related approach has been the use of reagents that cleave rRNA. The cleavage patterns reveal parts of the rRNA that are either buried or exposed.

Neutron diffraction of ribosomal proteins has been used to estimate the arrangement of ribosomal components. By reconstituting pairs of deuterated proteins into the 30S subunit, distances between the protein pairs were generated to produce a model of protein arrangement in the small subunit. Using similar methods several pairwise distances have been generated for the large subunit as well but, with over twice as many proteins, the task is significantly more difficult.

Together the accumulated biochemical and structural data have generated several models of ribosome structure. Several high-resolution structures of individual ribosomal proteins or fragments of rRNA bound to proteins have also been determined. However, the piecewise reconstruction of structural data seemed unlikely to provide an adequate picture of the ribosome at the molecular level. Recently, however, structures of isolated subunits and the assembled ribosome have been determined at a resolution that is already impressive and will continue to improve.

D. Recent High-Resolution Structures

Neither electron microscopy nor X-ray crystallography seemed suited to provide a high-resolution structure of the ribosome, which was thought to be too small for the first and too large for the second method. Yet both have been pushed to the limit to provide significant new structures that are already explaining mechanistic features of protein synthesis.

Cryoelectron microscopy (cryo-EM) has produced remarkable images of the entire ribosome as a result of two major advances in methodology—one experimental, one computational. Samples for cryo-EM (in this case of the 70S *E. coli* ribosome) are frozen quickly, encased in vitreous ice which lacks the crystals that can distort macromolecular structure. Following electron microscopy of these samples, a computer program digitally averages and aligns tens of thousands of images from single ribosomes at different angles. The resulting reconstruction produces a three-dimensional image at much higher resolution than any of the individual two-dimensional micrographs.

One of the first features observed by cryo-EM was the amount of visible deep clefts and "holes." Intersubunit contacts are limited to a discrete number of "bridges" between large and small subunits. These now appear to be composed primarily of RNA. The relatively small intersubunit surface is consistent with the need for subunits to dissociate following protein synthesis to accept a new mRNA molecule.

In addition to unoccupied 70S ribosomes, cryo-EM reconstructions are currently available for ribosomes containing various ligands. For example, tRNAs have been observed in each of the three ribosomal binding sites (A, P, and E), and ribosome-bound EF-Tu has also been visualized. The variety of structures available at up to 11 Å resolution points out an advantage of cryo-EM over X-ray crystallography for ribosomal structure determination. Because sample preparation is rapid compared to

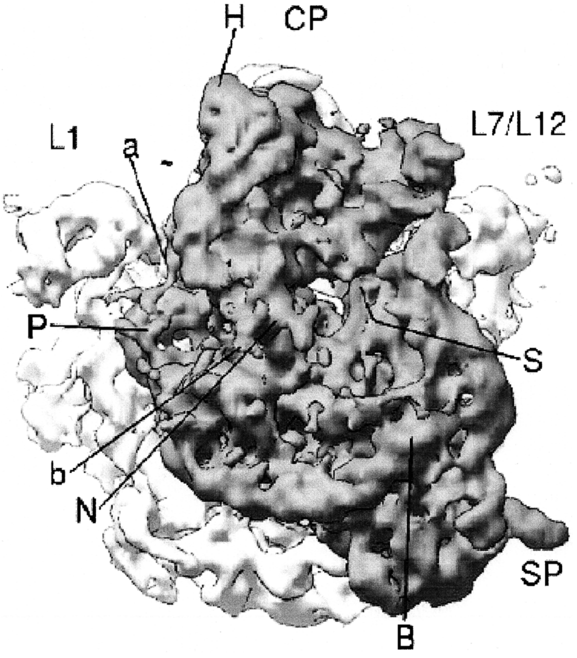


FIGURE 11 Bacterial ribosome structure at 7.8 Å. Architectural features of *Thermus thermophilus* 70S ribosomes are labeled, as identified by X-ray crystallography. The 30S subunit is the front, darker portion, and consists of the head (H) connected to the platform (P) and body (B). Other features of the 30S subunit are the neck (N), spur (SP), shoulder (S), and contacts between the head and platform (a and b). The 50S subunit includes the protein L1 stalk, central protuberance (CP), and L7/L12 region. [Reprinted with permission from Cate, J. A., Yusupov, M. M., Yusupova, G. Zh., Earnest, T. N., and Noller, H. F. (1999). "X-ray crystal structures of 70S ribosome functional complexes." *Science* **285,** 2095-2104. © 1999 American Association for the Advancement of Science.]

growth of crystals for X-ray diffraction studies, the many functional states of the ribosome can be more readily probed by EM to assemble a gallery of structures.

Even more surprising than the impressive cryo-EM structures have been recent leaps in resolution obtained by X-ray crystallography of ribosomes (Fig. 11). Although researchers have been able to generate diffracting crystals of ribosomes for years, only recently has the diffraction been of sufficient quality to identify clear regions of electron density corresponding to rRNA and protein. Here the lower-resolution cryo-EM structures have helped, in that they enabled crystallographers to position and orient the ribosomal particle within the crystal unit. Once the ribosome is properly oriented in the crystal, the heavy atom clusters used to determine phase angles can be located. Together these advancements provided remarkable images at moderate resolution of individual ribosomal subunits and the complete particle with mRNA and tRNAs bound. Atomic resolution structures will soon be forthcoming.

The structures now available at 5–8 Å resolution allow differentiation of protein α-helices from double-helical regions of rRNA based on their structural characteristics; some protein β-sheets can also be seen. The structures of several individual ribosomal proteins have been solved by X-ray crystallography or nuclear magnetic resonance spectroscopy, and the X-ray data on the whole 50S particle have been correlated with the previously determined structures of individual proteins. As expected, the accumulated data from IEM, cross-linking, and neutron diffraction helped to position these protein structures within the ribosome electron density and will continue to direct placement of proteins and specific regions of rRNA.

Although the locations of many proteins and thousands of nucleotides must still be teased out of the electron density maps, structural features of specific regions of the ribosome have already provided mechanistic clues. A more detailed understanding of the mechanism of protein synthesis will emerge from these structural studies.

E. Mechanistic Clues from Structure

On the interface side of the 30S subunit (where it interacts with the 50S subunit), a large region of rRNA is visible which lacks any protein density. This observation supports the theory that a primitive ribosome was composed only of RNA. Localization of the proteins of known structure revealed that S5 makes close contacts with S4; these were previously known to be near each other based on cross-linking and neutron diffraction data. Mutations in S4 and S5 produce error-prone ribosomes, and in fact these mutations are mapped to the area of contact between the two proteins. It seems likely, therefore, that inaccuracy results from a disruption of this protein–protein interaction.

A major feature observed on the 50S subunit structure is an apparent tunnel through the center of the subunit, through which the newly synthesized polypeptide is proposed to exit the ribosome. The tunnel seems to be wide enough to accommodate the polypeptide, and extends from the peptidyl transferase center to A-site on the back (solvent side) of the 50S subunit.

The crystal structure of the *Thermus thermophilus* 70S ribosome was solved with RNA substrates in the A, P, and E-sites of the ribosome and with bound mRNA. Interestingly, a partial tRNA mimic (an isolated anticodon stem-loop hairpin) bound in the P-site, held tightly in place by six contact points with the ribosome. These contacts likely stabilize the codon–anticodon interaction, properly orient the anticodon stem in the P-site, and may help position the A-site codon of the mRNA. The tRNA mimic bound in the A-site is not held tightly, but sits in a large cavity of electron density with only minimal contact points. This makes sense mechanistically, as the A-site serves to discriminate

between cognate and noncognate tRNAs, while the P-site must grip the peptidyl–tRNA to prevent dissociation from the ribosome.

F. Dynamics of Translation

Although the increasing resolution of cryo-EM and crystallographic structures promises molecular level information in the near future, these structures are snapshots of a dynamic particle. During each round of elongation, tRNA molecules move through three binding sites on the two ribosomal subunits, and the mRNA shifts by exactly one codon. While these movements are catalyzed *in vivo* by EF-Tu and EF-G and depend on GTP hydrolysis, they can also be observed *in vitro* in the absence of elongation factors. Whatever drives the molecular motor therefore must be intrinsic to the ribosome. In fact biochemical and genetic studies are beginning to identify switch mechanisms that may drive tRNA selection and translocation.

For example, the 16S nucleotides 885–890 and 910–912 can base pair in two different conformations (Fig. 12). The 910–912 CUC can form canonical Watson–Crick base pairs with the GAG of 888–890 or can pair with the GGG of 885–887 to give a central G–U wobble pair. Substitutions demonstrated that both pairings are essential for biological activity, suggesting that 16S rRNA switches between two conformations at different steps in protein synthesis. Furthermore, mutations that favored the "888 conformation" led to restrictive (hyper-accurate) ribosomal phenotypes, while those that stabilized the "885 pairing" produced error-prone phenotypes. Structurally, this triplet-switch-sequence has been located in the 70S crystal structure near the long 16S helix that is part of the decoding region of the small subunit. The switch helix is certainly involved in tRNA selection, and may also play a role in translocation, as suggested by the observation that mutations that favored the 888 conformation were hypersensitive to spectinomycin, an antibiotic known to inhibit EF-G-dependent translocation.

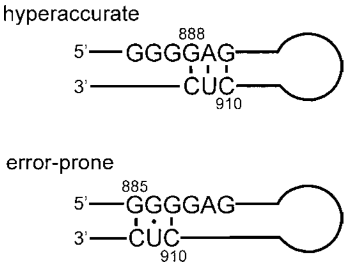


FIGURE 12 The 16S triplet switch. Genetic studies revealed that the 885–912 regions of 16S rRNA alternate between base-pairing conformations, as shown; both are required for cell viability. Mutations that favor the 888 conformation produce ribosomes that are hyper-accurate in protein synthesis, while substitutions that stabilize the 885 conformation lead to error-prone phenotypes. It is likely that this switch is involved in regulating tRNA selection and translocation.

XII. INHIBITORS OF PROTEIN SYNTHESIS

Many compounds have been identified which inhibit different steps in protein synthesis, either by mimicking translational substrates, blocking movement within the ribosome, or otherwise interfering with essential interactions (Table III). Some of these inhibitors are clinically relevant antibiotics and others have facilitated functional studies of the ribosome as described earlier. For example aminoglycosides such as streptomycin and gentamycin interfere with the decoding center on the 30S subunit. At low concentrations these antibiotics affect translational accuracy, while at increased doses they prevent formation of the 30S initiation complex, leading to cell death. Despite the increasing incidence of antibiotic resistance, the aminoglycosides remain useful therapeutic agents against gram-negative bacteria.

Puromycin is an antibiotic that mimics tyrosyl–or phenylalanyl–tRNA and binds in the ribosomal A-site (Fig. 13). The ribosome will even use puromycin as a substrate in one round of elongation, forming a peptide bond between the antibiotic and the P-site peptidyl–tRNA. This transpeptidation step terminates synthesis of the polypeptide and releases the abortive peptidyl–puromycin product.

XIII. POST-TRANSLATIONAL MODIFICATIONS

Many proteins, in order to perform their role in the cell, must be further processed after they are synthesized on the ribosome. These post-translational modifications involve

TABLE III Inhibitors of Protein Synthesis

Antibiotic	Activity
Chloramphenicol	Blocks peptidyl transferase function
Erythromycin	Blocks translocation of peptidyl–tRNA
Gentamycin	Prevents formation of 30S initiation complex
Kirromycin	Blocks release of EF-Tu:GDP from ribosome
Pseudomonic acid	Inhibits isoleucyl–tRNA synthetase
Puromycin	Mimics tyrosyl–tRNA, terminates elongation
Spectinomycin	Inhibits EF-G function
Streptomycin	Prevents transition from initiation to elongation
Tetracycline	Inhibits binding of aminoacyl–tRNAs to small subunit

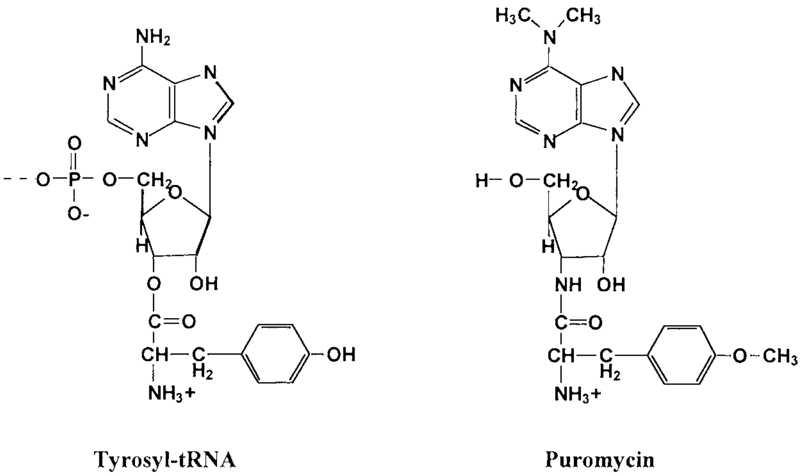


FIGURE 13 Puromycin—an AA–tRNA mimic. Puromycin (right) inhibits protein synthesis by binding in the ribosomal A-site as a mimic of tyrosyl–tRNA (left). Translation is terminated because puromycin does not have a free carboxyl group available for further elongation steps.

either proteolysis of the polypeptide backbone or chemical derivatization of protein functional groups.

Nearly all bacterial proteins are synthesized with a formyl–methionine as the first amino acid, as described above. Yet most proteins isolated from cell culture do not contain an N-terminal fMet or Met residue, indicating that the maturation process involves proteolytic removal of this amino acid. Some proteolytic events release a functional protein from a synthesized precursor, or proprotein. An example is insulin, which is released from its proinsulin precursor by excision of an internal 33-residue peptide. Another type of proteolytic processing is the removal of N-terminal signal peptides, which target some proteins for insertion into membranes or for secretion from the cell.

Proteins are also covalently modified by specific enzymes that act on side-chain functional groups or on the N- or C-termini. More than 150 types of side-chain modifications are known; these include glycosylations, methylations, and acetylations, among others. For instance, a variety of carbohydrates are added to proteins, primarily at asparagine residues, and serve as recognition markers on cell surfaces.

XIV. MITOCHONDRIAL PROTEIN SYNTHESIS

In addition to the different translational factors and mechanisms used by eukaryotic organisms as described earlier, an organizational difference exists in eukaryotes that contrasts with prokaryotic protein synthesis. In organisms lacking a nucleus, transcription of the genetic message from DNA to RNA occurs in the same location as translation. In fact, bacterial ribosomes typically begin translating an mRNA as it emerges from RNA polymerase. Eukaryotic protein synthesis is uncoupled from transcription, as mRNAs are synthesized and processed in the nucleus and then transported to the ribosomes in the cytoplasm.

While most protein synthesis occurs in the cytoplasm, mitochondria also have small chromosomes and synthesize some of their own proteins. For example, the human mitochondrial chromosome is composed of less than 17,000 bp and encodes 13 proteins, 22 tRNAs, and 2 rRNAs. Mitochondria use a genetic code that is slightly different from the "standard" code shown in Table I. Codon degeneracy is greater in mitochondria; for example, UGA is not used as a termination signal, but instead codes for tryptophan. Likewise, the trinucleotide AUA codes for methionine (and is also an initiation signal) in mammalian mitochondria, while in the standard code AUA specifies isoleucine. Last, only 22 tRNAs are used to decode all 61 sense codons, in contrast to the 60 or so that are used in prokaryotes or the cytoplasm of eukaryotes.

XV. NONRIBOSOMAL PEPTIDE SYNTHESIS

Nonribosomal peptides are a large family of naturally occurring compounds enzymatically synthesized from amino acids without a nucleic acid template. Many of these peptides have useful pharmacological properties. Vancomycin, for example, inhibits bacterial cell wall synthesis. Cyclosporin and precursors of penicillin are also nonribosomal peptides. These compounds are typically secreted by filamentous fungi, although their biological role in the producing organism is not known.

Nonribosomal peptide synthetases (NRPSs) are large modular proteins that assemble the amino acid units of the metabolites. Each module is responsible for a single catalytic step, such as activating an amino acid to its adenylate, forming a peptide bond, or modifying a functional group. The growing peptide is covalently attached to each module in turn, and the N- to C-terminal order of modules determines the organization of the final peptide product. For example, the penicillin precursor ACV is synthesized from modules specific for adenylate formation and attachment of amino adipate, cysteine, and valine, in that order. As more genes for NRPSs are identified, it should be possible to rationally construct novel nonribosomal peptides by genetic manipulation of modules.

XVI. PROTEIN SYNTHESIS AND THE RNA WORLD

In considering questions of the origins of life, many researchers have speculated that RNA was the earliest form of life, bridging the gap between a prebiotic era and the modern world of protein catalysts and DNA information storage. This suggestion initially came from the discovery of naturally occurring catalytic RNAs such as the *Tetrahymena* ribozyme and the RNA component of a tRNA processing enzyme known as RNase P. Even in extant systems, many enzyme cofactors have nucleotide or nucleotide-like components, such as ATP, NADH, and FAD. Furthermore, RNA both carries genetic information (mRNA) and performs catalysis (ribozymes).

If the "RNA world" hypothesis is correct, what evidence is there that early protein synthesis mechanisms were RNA controlled? First, there are three types of RNA that play major roles in translation—mRNA, which contains the genetic information; tRNA, the "adaptor" molecule; and rRNA, which comprises about 2/3 the mass of the ribosome. Mechanistically, some of the catalytic steps of protein synthesis have been mimicked using only RNA. Although several researchers have tried to demonstrate unequivocally that rRNA of modern organisms is the sole contributor of ribosomal function, the results are at best suggestive. However, the potential for RNA-catalyzed peptide bond formation is clear. In Darwinian-like experiments that select functional RNAs from large populations of partially randomized sequences, researchers have identified ribozymes that hydrolyze an aminoacyl ester, aminoacylate a tRNA, and form an amide bond between two amino acids. For example, such a selected ribozyme is able to catalyze formation of an amide bond between an AA–tRNA mimic and a peptidyl–tRNA mimic. Interestingly, this selected peptidyl transferase ribozyme contains some sequence and secondary structure elements of the region of *E. coli* 23S rRNA responsible for peptide bond formation.

SEE ALSO THE FOLLOWING ARTICLES

Hydrogen Bond • Macromolecules, Structure • Protein Structure • Ribozymes • Translation of RNA to Protein

BIBLIOGRAPHY

Buckingham, R. H., Grentzmann, G., and Kisselev, L. (1997). "Polypeptide chain release factors," *Mol. Microbiol.* **24,** 449–456.

Dever, T. E. (1999). "Translation initiation: Adept at adapting," *Trends Biochem. Sci.* **24,** 398–403.

Ibba, M., and Söll, D. (1999). "Quality control mechanisms during translation," *Science* **286,** 1893–1897.

Martinis, S. A., Plateau, P., Cavarelli, J., and Florentz, C. (1999). "Aminoacyl-tRNA synthetases: A family of expanding functions," *EMBO J.* **18,** 4591–4596.

Musier-Forsyth, K., and Schimmel, P. (1999). "Atomic determinants for aminoacylation of RNA minihelices and relationship to genetic code," *Acc. Chem. Res.* **32,** 368–375.

Nissen, P., Kjeldgaard, M., and Nyborg, J. (2000). "Macromolecular mimicry," *EMBO J.* **19,** 489–495.

Pennisi, E. (1999). "The race to the ribosome structure," *Science* **285,** 2048–2051.

Yarus, M. (1999). "Boundaries for an RNA world," *Curr. Opin. Chem. Biol.* **3,** 260–267.

Vitamins and Coenzymes

David E. Metzler
Iowa State University

I. The Discovery of Vitamins and Coenzymes
II. Nutritional Recommendations
III. Chemical Properties and Functions

GLOSSARY

Coenzymes Small organic molecules that function together with enzymes to catalyze specific types of chemical reactions.

Growth factor A material needed in small amounts for growth of an organism. Growth factors for animals are vitamins but other substances may also be growth factors for bacteria, yeasts, etc.

Prosthetic groups Coenzymes or similar cocatalytic groups that are firmly attached, usually by covalent linkages, to enzymes.

Vitamins Naturally occurring organic compounds that for human beings are essential nutrients in very small amounts. Vitamins are often components of coenzymes and prosthetic groups.

COENZYMES are small organic molecules that function with thousands of different enzymes in all organisms, assisting in the catalytic processes needed for life. They often contain vitamins as components. Several coenzymes participate in the major oxidation and reduction processes of cells. Some assist in the making and breaking of carbon–carbon bonds. Others are carriers of molecular fragments. Coenzymes participate in virtually every aspect of the chemistry of every living cell whether of bacteria, protozoa, fungi, higher plants, animals, or human beings. Some vitamins are incorporated into the coenzymes in which they function. Others become chemically attached to proteins and act as bound coenzymes, often described as prosthetic groups. Vitamin C (ascorbic acid) functions as a coenzyme without any firm attachment to a protein. Vitamin A, in one of its forms (retinoic acid), acts as a hormone. Vitamin D, which can be formed in the skin by the action of sunlight, is a natural precursor to oxygen-containing derivatives which also act as hormones. Most modern human beings would have difficulty in meeting the body's needs for vitamin D by sunbathing; hence the designation of this prohormone as vitamin D.

I. THE DISCOVERY OF VITAMINS AND COENZYMES

As early as 1750 it was recognized that green vegetables and citrus fruits could prevent the dread disease scurvy, which afflicted ancient sea voyagers, causing hemorrhages of skin, gums, and joints, followed by death. At about that time, Captain James Cook showed that sailors could avoid scurvy during long voyages by eating local green vegetables and grasses. Soon thereafter British seamen became "limeys." The chemical structure of the active component, vitamin C (Fig. 1), was established in 1933. In a similar

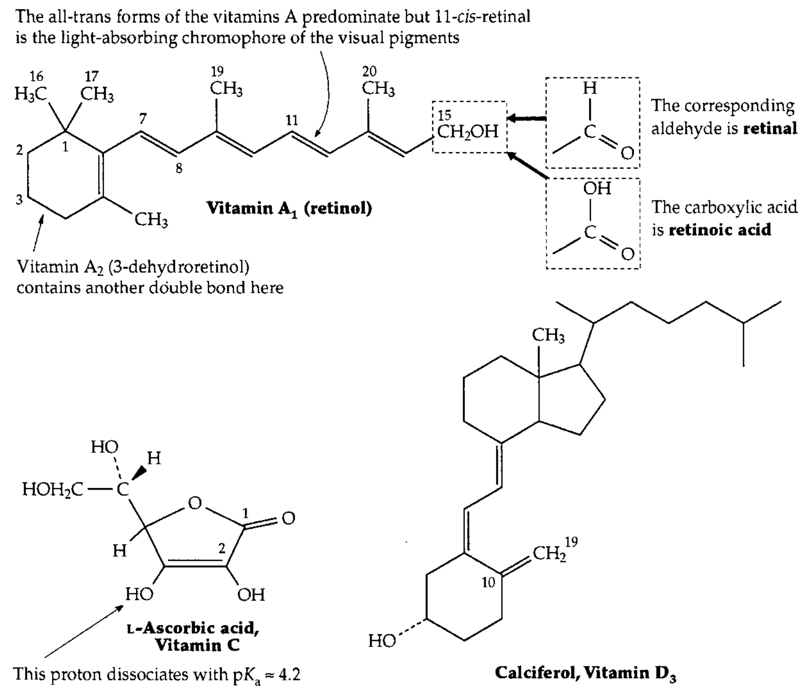


FIGURE 1 The structures of vitamins A, C, and D.

way, from about 1830 cod liver oil was used to prevent rickets. The active ingredient was vitamin D (Fig. 1). Later, vitamin A (Fig. 1), also present in cod liver oil, was recognized for its prevention of night blindness and maintenance of healthy skin. In oriental countries the disease beriberi, with its strange paralysis called polyneuritis, killed millions. A persuasive demonstration that this was a deficiency disease came in 1893 when Eijkman, working in Indonesia, demonstrated that chicks fed the white rice consumed by the local populace developed a rapidly fatal paralysis. However, the chicks could be completely cured by prompt feeding of a rice bran extract. It was 1926 before the curative compound was isolated from rice bran, characterized chemically, and named thiamin (Fig. 2). In 1912 the Polish biochemist Casimir Funk proposed that the four diseases scurvy, beriberi, pellagra, and rickets resulted from the dietary deficiency of vital nutrients which he imagined to be amines. He called them vitamines.

At about the same time, McCollum and Davis and others discovered that rats fed on semi-artificial diets required small amounts of "accessory growth factors." Growth of rats required both a fat-soluble material A and a water-soluble material B. Factor A, which could be found in milk, was later shown to consist of what we now call vitamins A, D, and E (Fig. 3). In 1939 another essential fat-soluble nutrient, vitamin K (Fig. 3), was isolated from plant sources. It was designated K for koagulation, because it was needed for blood clotting. The water-soluble factor B cured beriberi in chicks. However, it was also shown to consist of several components, and could better be described as a B complex. The beriberi curative factor thiamin, which was easily destroyed by heat, was designated B_1. Another nutritionally essential component, B_2, was stable to heat. The major growth-stimulating component of B_2, the yellow fluorescent compound riboflavin (Fig. 4), was designated vitamin B_2. Other water-soluble components were identified using a variety of tests. Nicotinamide was found in the coenzyme nicotinamide adenine dinucleotide (NAD; see Fig. 8) in 1935. The corresponding carboxylic acid nicotinic acid (also called niacin, Fig. 4) is also an active vitamin. It was a well-known compound

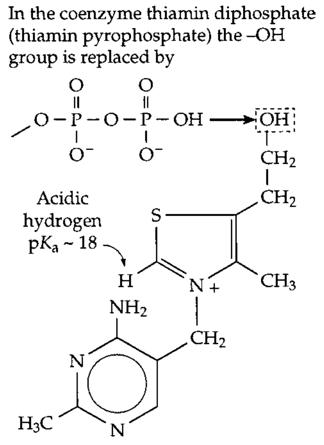


FIGURE 2 The vitamin thiamin and its coenzyme form thiamin diphosphate (thiamin pyrophosphate).

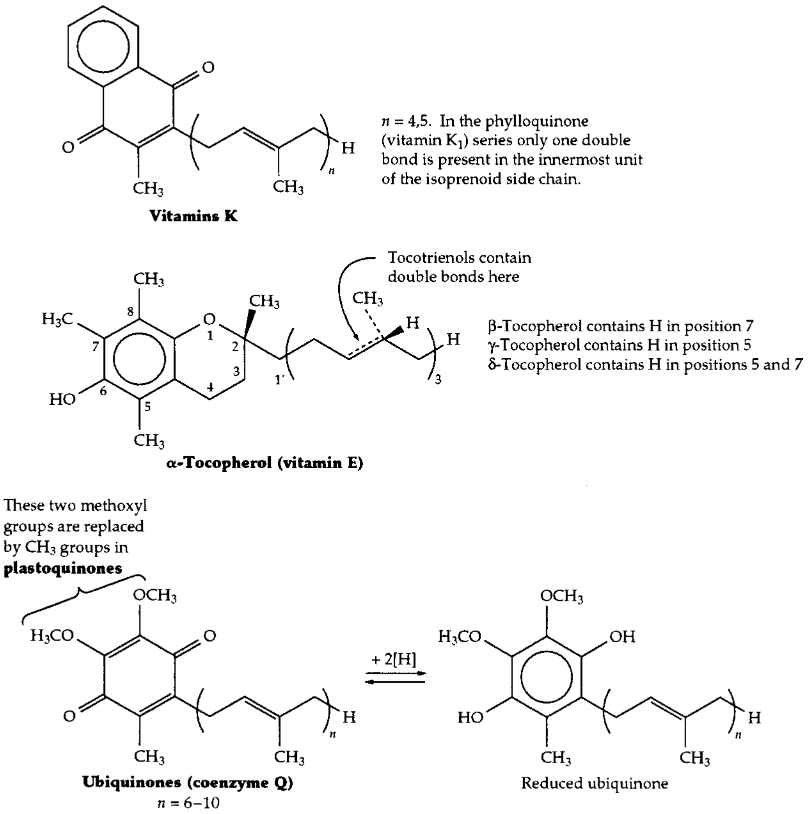


FIGURE 3 Structures of vitamins K and E and of the related ubiquinone (coenzyme Q) and plastoquinone.

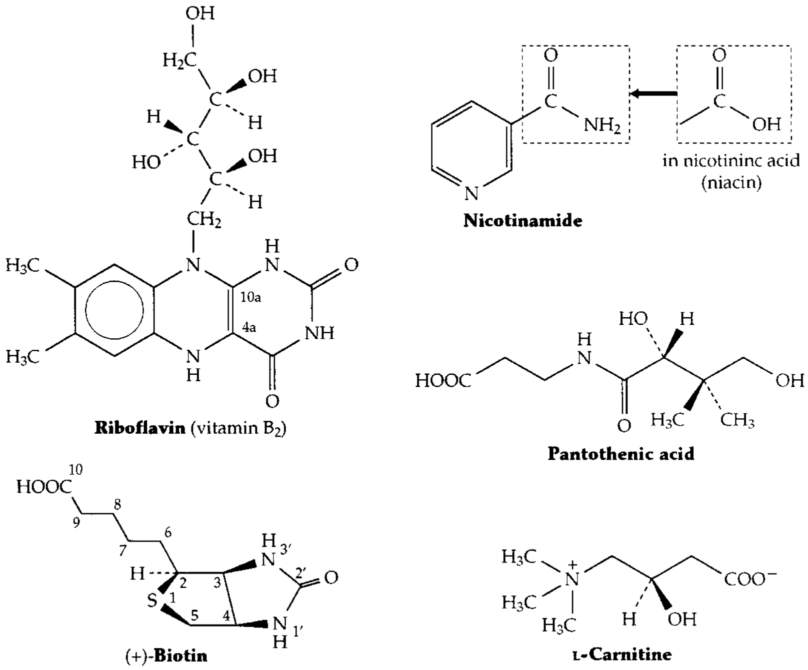
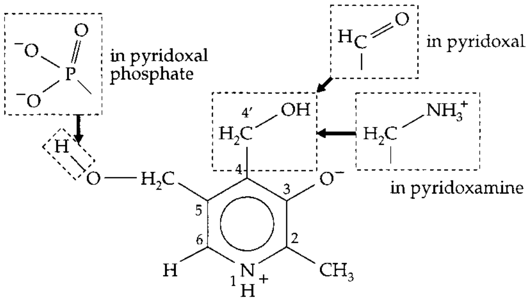


FIGURE 4 Four components of the vitamin B complex: riboflavin, nicotinamide, biotin, pantothenic acid, and the acyl carrier carnitine. Thiamin (vitamin B_1, Fig. 2) is also a member of the B complex. Carnitine, which has an essential acyl carrier function in the human body, is a vitamin for flour beetles.

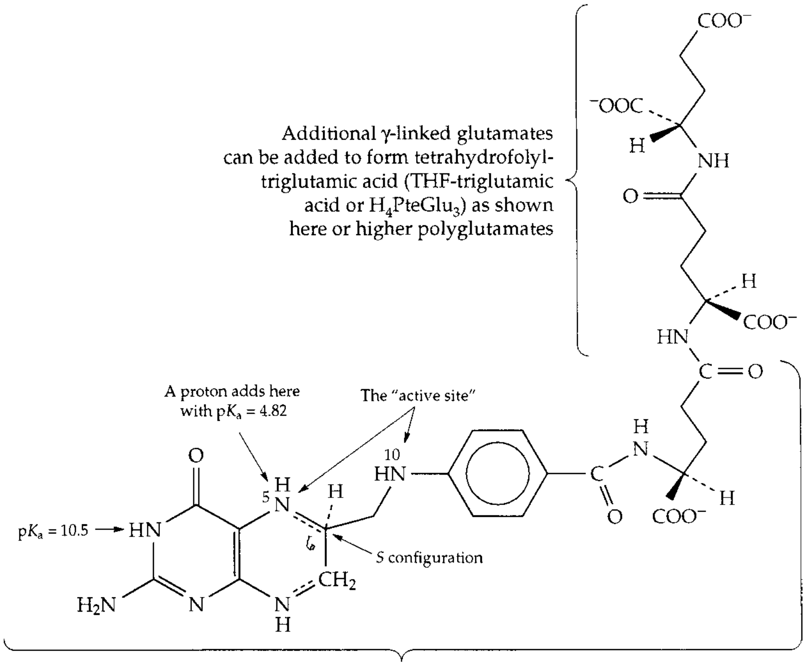


Pyridoxine (pyridoxol)
Vitamin B_6 alcohol

FIGURE 5 The vitamin B_6 (pyridoxine) family: pyridoxol, pyridoxal, pyridoxamine, and the coenzyme forms pyridoxal and pyridoxamine phosphates.

that had been prepared in 1867 by oxidation of nicotine, but had not been recognized as a nutrient until 1935 when it was shown to cure "black tongue" of dogs and shortly thereafter human pellagra. Biotin (Fig. 4) was first identified as a growth factor for yeast. Pantothenic acid (Fig. 4), which is present in plentiful amounts in many foods, was recognized as a curative agent for a dermatitis of chicks. Vitamin B_6 (Fig. 5) prevented a facial dermatitis in rats ("rat pellagra"). Folic acid, first described by Lucy Wills in Bombay, was identified as a vitamin that cured a macrocytic anemia that was often associated with pregnancy. The curative material, which is abundant in green leafy vegetables, was named folic acid. However, this name is usually reserved for the synthetic compound used in dietary supplementation. The natural forms are largely the coenzymes (Fig. 6), which are collectively called folates. The last of the accepted human vitamins to be discovered was vitamin B_{12}. A cobalt-containing organic compound needed in very small amounts, it cures and prevents pernicious anemia, which was often a fatal disease of people over 60 years of age. Its complex structure (Fig. 7) was determined by X-ray diffraction after numerous efforts at chemical characterization had failed. However, cyanocobalamin, the compound isolated and the form used in nutritional supplementation, is an artifact of the isolation and synthesis. The natural vitamin may have OH in place of CN but consists largely of the coenzyme forms.

Have all of the vitamins been discovered? Rodents have been reared on almost completely synthetic diets. However, good health in human beings may require additional materials. For example, some essential compounds might be made by intestinal bacteria. Some essential coenzymes such as lipoic acid, ubiquinone (coenzyme Q), and pyrroloquinoline quinone (PQQ) may be vitamin-like. Their presence in foods may be beneficial. Some individuals may need as dietary components these compounds, which are normally made by the body. Another example is the acyl-carrier molecule carnitine (Fig. 4), which



Tetrahydrofolic acid (THF) or tetrahydropteroylglutamic acid ($H_4PteGlu$)

FIGURE 6 The coenzyme tetrahydrofolic acid (tetrahydropteroylglutamic acid). The vitamin folic acid has two additional double bonds (dashed) in the second ring. Most of this coenzyme exists within cells as more complex forms containing additional glutamic acid units attached to the side chain at the upper right.

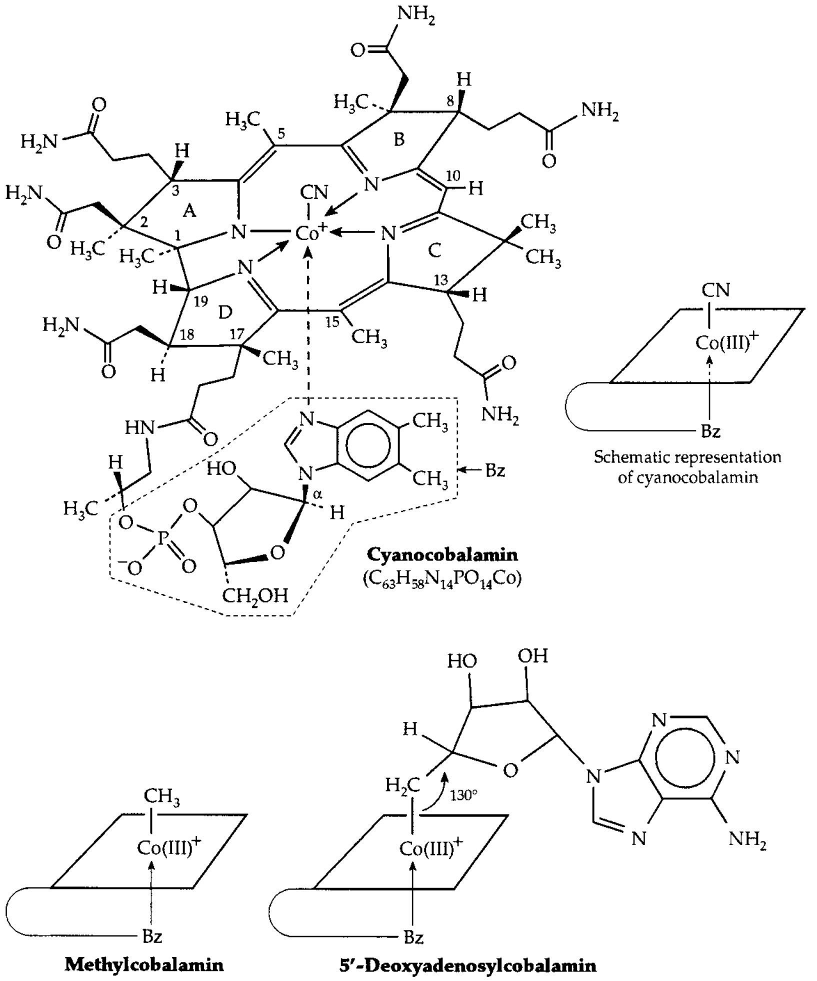


FIGURE 7 Vitamin B_{12} (cobalamin), the first discovered carbon-cobalt compound and its coenzyme forms methylcobalamin and 5′-deoxyribosylcobalamin.

is a growth factor needed by a common species of flour beetle, but is synthesized in adequate amounts by most human individuals. However, a few children require dietary carnitine. Inositol is an essential growth factor for yeast and is sometimes regarded as a vitamin. Flavonoid compounds, abundant in citrus fruit and other plant sources, have also sometimes been classified as a vitamin. Several amino acids, common constituents of proteins, must also be present in the human diet. These are needed in relatively large amounts and are not classified as vitamins. The role of small amounts of another amino acid taurine, which is essential for cats, in human nutrition is of current interest. The essential omega-3 and omega-6 fatty acids are also necessary dietary constituents, as are numerous metallic elements.

The discovery of coenzymes and catalytic prosthetic groups came in part from biochemical studies of yeast, of alcoholic fermentation, and of respiration. Buchner in 1899, discovered that a cell-free juice prepared from freshly ground active yeast cells still fermented sugar. Pasteur and others had previously maintained that fermentation required intact cells. Dialysis of the yeast juice stopped the fermentation, but the material that diffused out, which was called cozymase, could be added back with restoration of the fermentation ability of the juice. Cozymase was soon found to consist of a mixture of the nicotinamide-containing compound now called NAD (Fig. 8), magnesium ions, and thiamin diphosphate (Fig. 2). NAD was quickly recognized as a component of a "respiratory chain" in animal and yeast cells. Since about 1910 this chain has been recognized as beginning with the hemoglobin-like oxygen-binding protein cytochrome oxidase and an intensely yellow flavin compound that was identified as the riboflavin derivative FAD (Fig. 9). The cooperative functioning of these vitamin-containing compounds, together with associated proteins, established the concept of coenzymes. Additional compounds, isolated from natural materials, containing vitamin B_6, pantothenic

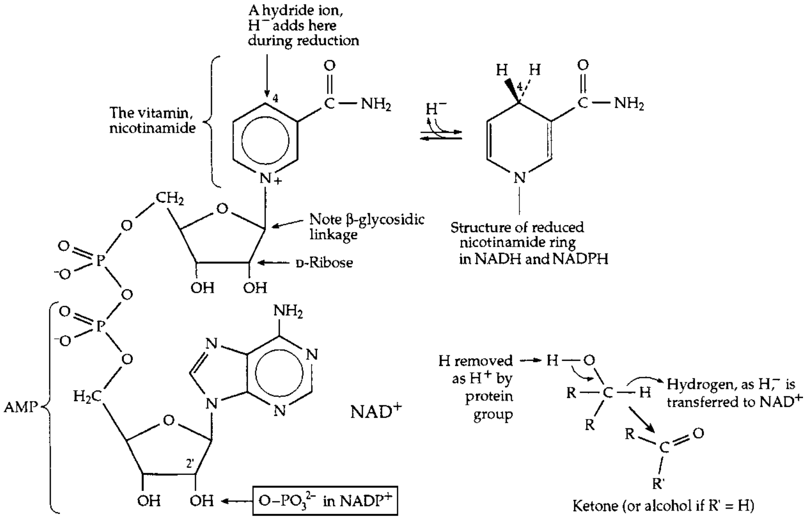


FIGURE 8 The nicotinamide-containing coenzymes nicotinamide-adenine dinucleotide (NAD) and nicotinamide-adenine dinucleotide phosphate (NADP). Also illustrated are their biological functions as hydrogen carriers. The nicotinamide ring accepts a hydride ion (H^-) transferred directly from a substrate molecule.

acid (coenzyme A, Fig. 10), folic acid, and vitamin B_{12} are among many substances that are now described as coenzymes.

NAD, NADP, and thiamin diphosphate were found to bind reversibly to their host proteins. NAD and NADP, as their reduced (NADH, NADPH) and oxidized (NAD^+, $NADP^+$) forms (Fig. 8), were found to act as hydrogen carriers, moving freely between two or more catalytic proteins. In contrast, FAD and pyridoxal phosphate (PLP, Fig. 5) are extremely tightly bound to some proteins and normally function without dissociation from the catalytic protein. Still others, such as biotin, are covalently bonded to proteins (Fig. 11). The same is true of some FAD derivatives. These tightly bound cocatalysts are often referred to as prosthetic groups. These include a great variety of both organic and metallo-organic structures. Among the latter are the heme proteins. Vitamin C (ascorbic acid or ascorbate; Fig. 1) is unusual in functioning largely in a free, unbound form, and often at a very high concentration. This is also consistent with its high nutritional requirement for human beings. Vitamin A has a special role in vision. The aldehyde retinal (Fig. 1) combines with proteins of the retina to form the light receptors of the visual cells. Vitamin K has a specialized function in formation of a series of proteins needed for blood clotting. Both vitamin A (as retinoic acid) and vitamin D (as hydroxylated derivatives) serve as important hormones.

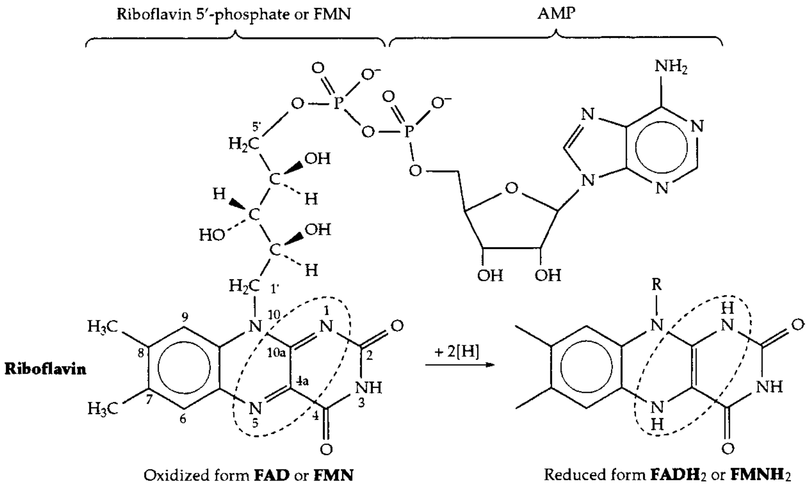


FIGURE 9 The coenzyme forms of riboflavin, riboflavin 5′-phosphate (FMN) and flavin-adenine dinucleotide (FAD).

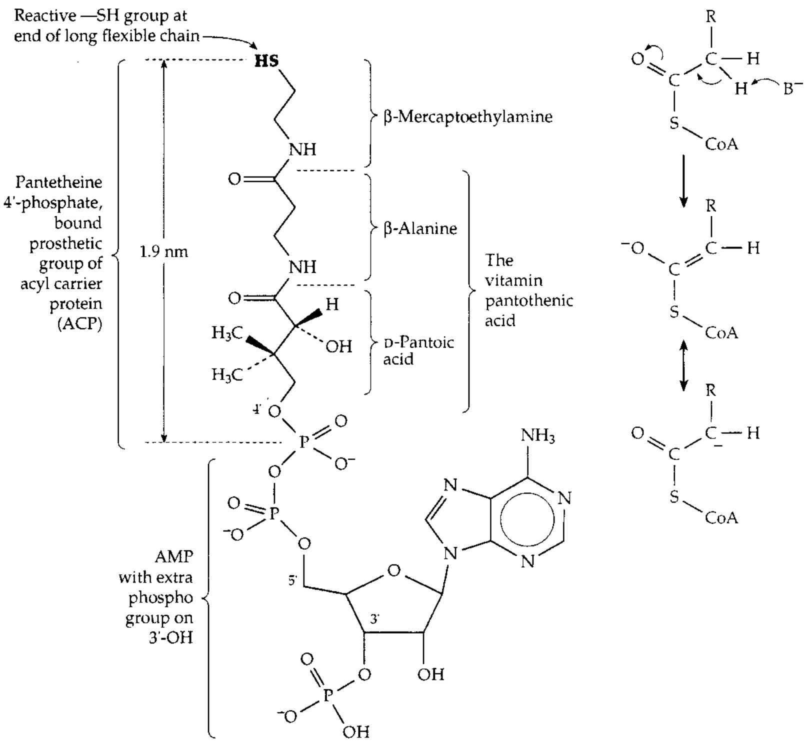


FIGURE 10 Coenzyme A and its constituent components, which include the vitamin pantothenic acid.

II. NUTRITIONAL RECOMMENDATIONS

It is difficult to establish the amount of any vitamin that is essential to a human being. Even for animals the amount required for good health must exceed that needed for survival. The enormous individual variation among human beings ensures that any conclusion about the requirement for a nutrient will be incorrect for at least some individuals. The need for a vitamin will be affected by the age of the individual, by differences in the ability to take up the vitamin from the digestive tract, and by the ability to convert it to appropriate coenzyme forms. Also important is the ability of numerous enzymes to hold the coenzyme correctly into their active sites. With many thousands of possible sites for mutation of the DNA encoding these proteins, there are many possible reasons why some individuals may need larger amounts of a vitamin than does the average person. Recommended dietary allowances (Table I) are based on studies by panels of nutritional investigators. They vary somewhat from one country to another and are revised periodically. To put the needs for vitamins in a better perspective, Table II lists in concise form the other known human nutritional requirements.

Most of the B vitamins are synthesized by plants, fungi, and bacteria. Meat and dairy products also contain vitamins that have been obtained from these sources. As a consequence, a well-balanced human diet usually supplies adequate amounts of all of the vitamins. There are exceptions. Vitamin B_{12} is not made by plants and strict vegetarians may become deficient if their diet does not contain yogurt or other products of fermentation. Thiamin is very labile, especially at high pH. Cooking at a pH above 8 quickly destroys this vitamin. Alcoholism is another cause of thiamin deficiency, sometimes leading to the characteristic Wernicke's disease or Wernicke–Korsakoff syndrome, conditions with specific symptoms of encephalopathy. Because of the lability of thiamin, the number of turnovers, i.e., the number of times that a thiamin diphosphate molecule can pass through its catalytic cycle, seems to be limited. For this reason, the recommended daily allowance is increased by 0.5 mg for each 1000 kcal (Cal) consumed beyond that for an average person. Riboflavin is destroyed by light. Diets in many parts of the world are deficient in vitamin A and also in the plant carotenes, which can be converted into vitamin A in the body. Folate deficiency may occur if there is inadequate intake of fresh vegetables and fruit. Prolonged cooking can also destroy the vitamin.

Specific dietary deficiencies sometimes affect large populations. In the past, beriberi was a widespread consequence of consumption of polished rice without vitamin supplementation. During the early decades of the last century, pellagra was widespread in southern regions of the United States because the diet was low in protein

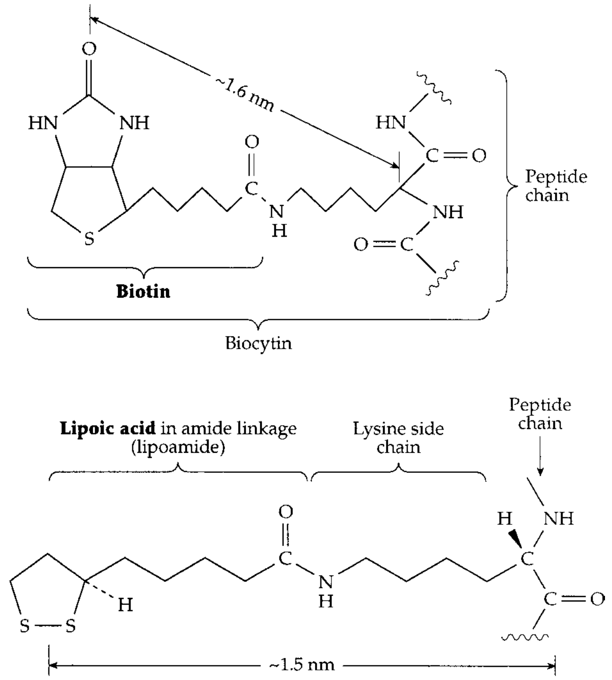


FIGURE 11 The vitamin biotin and the vitamin-like compound lipoic acid and their covalent attachments to selected lysine side chains in proteins (polypeptides). Both of these compounds function as catalytic prosthetic groups, biotin for CO_2 and lipoic acid for hydrogen. The fragment biocytin was isolated from autolysates of rapidly growing yeast.

and high in maize, a grain whose protein is deficient in tryptophan. Tryptophan can be converted to nicotinamide with an efficiency of about 1/60. Hence, most diets provide the necessary minimum. However, persons with pellagra often died after suffering from characteristic symptoms of dermatitis, diarrhea, and dementia. Deficiency of vitamin D was widespread, especially in northern regions, prior to the use of supplementation of milk. Deficiencies of the B vitamins, pantothenic acid, riboflavin, biotin, and vitamin B_6, are not often met in the human population. Except for the sensitivity of riboflavin to light, these compounds are quite stable. Nevertheless, some infants are born with unusually high requirements for specific vitamins. Some cases of sudden infant death have been attributed to biotin deficiency and convulsions in infants to a deficiency of vitamin B_6 in a nutritional formula. Vitamin B_6 is a family of three forms, an alcohol pyridoxol, an amine pyridoxamine, and an aldehyde pyridoxal (Fig. 5). Of these, pyridoxol, a very stable compound, predominates in plants. More of the vitamin is present as the less stable pyridoxal and pyridoxamine in foods of animal origin.

Vitamin C is made not only by plants but also by most animals who use the sugar glucose as the starting material. However, human beings, guinea pigs, and a few other

TABLE I Approximate Nutritional Requirements (mg/day) for the Vitamins and Some Characteristic Deficiency Diseases or Symptoms

Vitamin	Approximate daily need (mg)	Deficiency diseases	Related coenzyme or function
Thiamin	0.8 or more[a]	Beriberi	Thiamin diphosphate
Pantothenic acid	10–15		Coenzyme A
Riboflavin	1.5		FMN, FAD
Nicotinamide (or nicotinic acid)	2.5[b]	Pellagra	NAD, NADP
Biotin	0.15–0.3		Bound as prosthetic group
Pyridoxine (vitamin B_6) phosphate	1.5–2		Pyridoxal or pyridoxamine
Folic acid	0.2–0.4[c]		Tetrahydrofolate
Vitamin C	50–200	Scurvy	Antioxidant, electron carrier
Vitamin B_{12} (cobalamin)	0.002	Pernicious anemia	5′-Deoxycobalamin, 5′-methylcobalamin
Vitamin A (retinol)	0.7		Retinol, bound as prosthetic group
Vitamin D	0.02	Rickets	Hormonal role in calcium metabolism
Vitamin E	8–10		Antioxidant
Vitamin K	0.05–0.08	Bleeding	Blood clotting

[a] Amount should be at least 0.5 mg per 1000 kcal (Cal) of food energy.
[b] Some may be obtained from metabolism of the amino acid tryptophan, about 1/60 of which can be converted into this vitamin.
[c] The larger amount is recommended for women of child-bearing age.

species are unable to synthesize this important antioxidant compound. The need for ascorbic acid is high, but the optimum amount needed for good nutrition is uncertain. Furthermore, there has been some concern that excessive intake of vitamin C, especially in combination with iron ions, may generate damaging free radicals. However, ascorbic acid seems to have predominantly an antioxidative effect in animals.

Vitamin B_{12} is required in minute amounts, one microgram per day supplying the needs for the human body. However, absorption of this small amount of vitamin from the gut and transport to its sites of action requires special transport proteins. One of these, the "intrinsic factor," is synthesized by cells of the intestinal mucosa and is utilized for absorption of vitamin B_{12}. Synthesis of the intrinsic factor is defective in some individuals, and is often inadequate in persons older than about 60 years. If untreated, this deficiency leads to pernicious anemia, a

TABLE II Other Human Nutritional Requirements and Some Biological Functions[a]

Nutrient	Approximate daily need (mg or g)		Major biological function
Water	Variable		Solvent
Energy	A. Basal need ~1800 kcal (Cal)/day		Metabolism
	B. Additional needed for work: 240 g carbohydrate or 108 g fat per 1000 additional kilocalories (Cal)		
Major energy sources			
Carbohydrates (4.1 kcal/g)	300 g*	230 kcal (Cal)	*These amounts together will supply typical basal need
Fat (9.3 kcal/g)	65 g*	605 kcal (Cal)	
Protein (4.1 kcal/g)			
Protein for biosynthesis	~0.44g/kg body weight (for 70 kg person, 31 g)		Must include the nine essential amino acids plus 11 other amino acids needed for protein synthesis and other purposes or other suitable nitrogen source for their synthesis.
Essential amino acids			All of these, as well as the "nonessential" amino acids, are needed for formation of specific proteins in the body. Several are also required for synthesis of nucleotides, coenzymes, hormones, and neurotransmitters.
	Infants	Adults (older)	
Valine	93	20 (10)	
Leucine	160	39 (14)	
Isoleucine	70	23 (10)	
Methionine (+cysteine)[b]	58	15 (13)	
Phenylalanine (+tyrosine)[b]	125	39 (14)	
Tryptophan	17	6 (4)	
Threonine	87	15 (7)	
Lysine	103	30 (12)	
Histidine	28	8–12	
Essential fatty acids			Absolutely required. Enter cell membranes and affect many biochemical processes. The C20 acids are also converted to eicosanoids, signaling molecules that include prostaglandins and leukotrienes. Essential fatty acids protect against cardiovascular disease, disease, inflammation, and autoimmune reactions.
Omega 6 (ω6 or n-6)	1–4 % of total calories		
Linoleic acid (C 18:2, 18 carbon atoms, 2 *cis* double bonds) and arachidonic acid (C 20:4)			
Omega 3 (ω3 or n-3)	0.1–0.3% of total calories		
Linolenic acid (C18:3), eicosapentaenoic (C20:5), and docosahexaenoic acid (C22:6) acids			
Mineral elements			
	Infants	Adults	
Sodium Na^+			Electrolyte
Potassium K^+			Electrolyte
Chlorine Cl^-			Electrolyte
Calcium Ca^{2+}	270	1000	Structural in proteins, carbohydrates, bone; signaling ion

continues

TABLE II (*Continued*)

Nutrient	Approximate daily need (mg or g)		Major biological function
Phosphorus P	275	700	Present in nucleic acids, proteins, coenzymes
Magnesium as Mg^{2+}	75	300	Enzyme activator, often associated with organic phosphate groups; electrolyte
Zinc as Zn^{2+}	5	15	Structural; catalytic component in active sites of enzymes
Iron Fe	1	1 (men) 2 (young women)	Active sites of oxidative enzymes, electron transport proteins
Copper Cu	1.5–3 mg		Oxidative enzymes, electron-transferring proteins
Manganese Mn	2–5 mg		Component of enzymes
Iodine I	150 μg		Formation of thyroxine, triiodothyronine
Sulfur S			Largely supplied as cysteine or methionine (above)
Selenium Se	50 μg		Formation of selenocysteine, component of active sites of several enzymes and other proteins
Molybdenum Mo	25 μg		Formation of sulfite oxidase and other molybdoenzymes
Chromium Cr	50 μg		Utilization of glucose
Cobalt Co	as vitamin B_{12} (Table I)		
Ultratrace elements, probably needed or beneficial			Most functions are uncertain
Boron B	1–10 mg		Crosslinking?
Fluorine F	1.5–4 mg		Protective component of hydroxyapatite in teeth, bones
Arsenic As	15 μg		
Silicon Si	5–30 μg		Crosslinking in connective tissue
Nickel Ni	25–35 μg		Uncertain
Vanadium V			Component of thyroid peroxidase
Possibly needed	Typical dietary intake		Functions are unknown
Aluminium Al	2 mg		
Bromine Br	2–8 mg		
Cadmium Cd	0–20 μg (toxic in excess)		
Germanium Ge	0.4–1.5 mg (toxic in excess)		
Lead Pb	15–100 μg (toxic in excess)		
Lithium Li	0.2–0.6 mg		
Rubidium Rb	1–5 mg		
Tin Sn	1–40 mg		

[a] Data are from Shils, M. E., *et al*., eds. (1999). *Modern Nutrition in Health and Disease*, 9th ed., Williams & Wilkins, Baltimore. This book can be consulted for detailed discussions of all of the listed dietary components.

[b] The need for methionine is decreased if cysteine (or cystine) is present. Likewise, tyrosine decreases the need for phenylalanine. Persons with phenylketonuria must have tyrosine.

condition in which red blood cells do not mature normally and in which dementia develops as a result of the lack of vitamin B_{12} in the brain. If treated in time, a monthly injection of one milligram of the vitamin is curative.

A deficit of vitamin A causes night blindness and loss of proper differentiation of epithelial cells. A dangerous symptom is the dry eye condition xerophthalmia, which can cause blindness. In fact, thousands of children in developing countries become blind from this condition each year. Fortunately, the problem can be alleviated inexpensively. A single oral dose of vitamin A provides a store in the liver adequate for 4–6 months. An international effort to eradicate vitamin A deficiency as a cause of blindness is in progress. Deficiency also interferes with reproduction. The yellow beta-carotene and some related plant pigments can be converted by the human body into vitamin A. About six micrograms of all-*trans* beta-carotene yields one microgram of the vitamin. In large excess, vitamin A, especially in the form of retinoic acid, is toxic. About 3 mg

per day of retinol or naturally occurring retinol esters is a safe limit. Amounts of vitamin A are often given in international units (IU). One IU is provided by 0.3 μg of all-*trans* retinol.

Deficiency of vitamin K is rare in adults but more frequent in breast-fed infants. The characteristic symptom of slow blood clotting may also arise, rarely, because of a hereditary lack of vitamin K-dependent processing of blood clotting proteins. The exact functions of vitamin E have been hard to define, but a deficiency can cause neurological and reproductive problems and muscular dystrophy in some animals. Although symptoms are rare in humans, they appear in various hereditary conditions such as the lack of a liver tocopherol transport protein. There are eight naturally occurring isomeric forms of vitamin E (Fig. 3) with differing potencies. The most active is the natural *R*, *R*, *R*-isomer of α-tocopherol for which 0.67 mg = 1 IU. At high levels, e.g., 1200 IU per day, vitamin E may compete with vitamin K and cause bleeding.

III. CHEMICAL PROPERTIES AND FUNCTIONS

The major chemical components of cells include the nucleic acids RNA and DNA, polysaccharides (carbohydrates), fatty materials (lipids), and many thousands of different proteins. Proteins catalyze most of the metabolism, the network of chemical reactions by which cells construct their own substance and by which they obtain and utilize energy for all life processes. Proteins, whose structure is dictated by the DNA of the corresponding genes, are precisely constructed. As submicroscopic machines they have moving parts and apparatus for recognizing and binding to other molecules, both large and small. Some coenzymes cooperate with the proteins by carrying electrons, atoms, or molecular fragments. Others help the proteins to catalyze reactions that are difficult or impossible for the reactive groups provided by the amino acid side chains of the proteins.

The B-vitamin-containing coenzymes provide a logical starting point for our discussion. As mentioned in Section I, thiamin, nicotinamide, and riboflavin were recognized early in the 20th century as participants in energy metabolism in both animal and yeast cells. Panthothenic acid, biotin, and vitamin B_{12} were soon added to this list. We now know, in part from complete genome sequences, that all living creatures depend upon these coenzymes to help catalyze a series of central pathways of metabolism. One of these pathways is utilized by aerobic organisms, from bacteria to human beings, for the oxidation of fatty acids (Fig. 12).

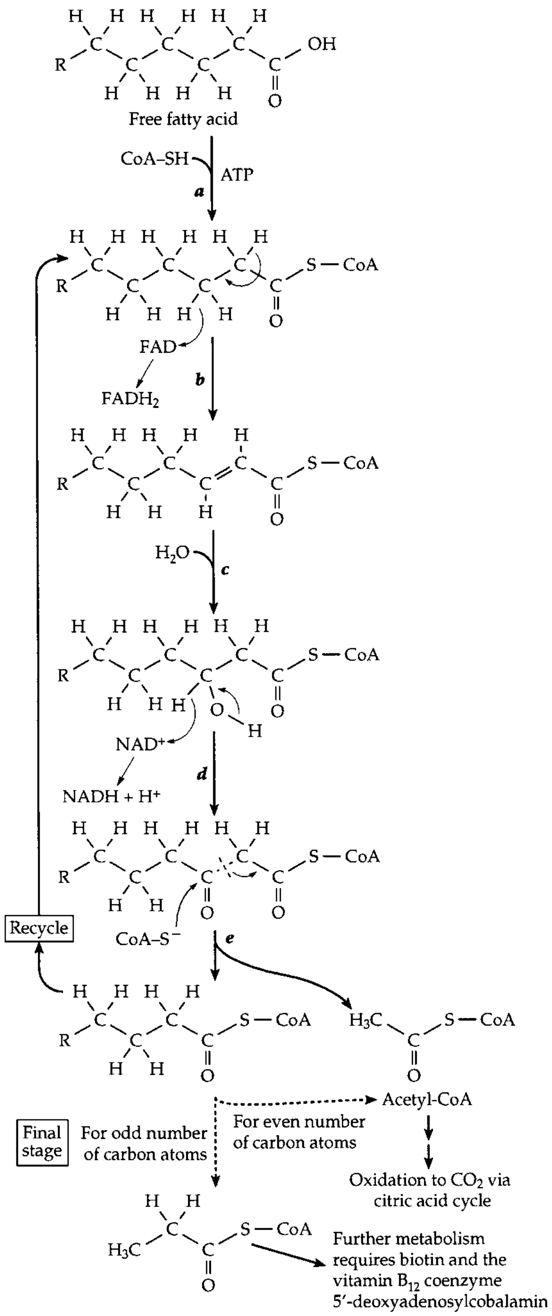


FIGURE 12 The use of five different vitamin-containing coenzymes in an important metabolic process, the oxidation of fatty acids (beta oxidation) to carbon dioxide and water.

A. Coenzyme A, an Acyl Group Carrier and Activator

Coenzyme A, named for its role as an acetyl group carrier, contains the vitamin pantothenic acid as an essential constituent (Fig. 10). The synthesis of this vitamin can be accomplished by green plants, fungi, and most bacteria, but not by the human body. Its unusual chemical structure

provides the necessary shape and chemical properties to allow it to bind into crevices within the active sites of a variety of proteins. There it not only fits snugly but has electrostatic bonding interactions that allow the proteins to hold it in just the correct orientation for its function. Curiously, the exact biological need for the unusual structure of the vitamin is still obscure. The chemically functional end of coenzyme A is the sulfhydryl (—SH) group which is added on to the vitamin structure by cells, as shown in Fig. 10. Before coenzyme A can function it must be combined in a thioester linkage with a carboxylic acid such as acetic acid, or a long-chain fatty acid, as illustrated in Fig. 10 (right). It is customary in discussions of metabolism to indicate the bulk of the coenzyme structure as CoA. The free coenzyme is designated CoA—SH, and in a thioester the hydrogen of the —SH group is replaced by an acyl group. The coenzyme has two functions. First, it can carry the acyl group from one protein to the next in a metabolic sequence, such as that of Fig. 12. Second, it can activate a hydrogen atom adjacent to the carbonyl (C=O) group for removal of a proton (H^+) by a catalytic group of basic nature, such as $-NH_2$, present in the protein. The carbonyl group in a thioester is an electron accepting group, whose facilitation of the proton removal is often indicated by curved arrows, as shown in Fig. 10 (right). The product of this proton removal is a reactive anion which is able to undergo formation or cleavage of carbon–carbon bonds or dehydrogenation by the riboflavin-containing FAD, as shown in Fig. 12.

Other coenzymes and prosthetic groups may also act as acyl group carriers. For the biosynthesis of fatty acids, a shortened version of coenzyme A (phosphopantetheine, Fig. 10), is covalently linked to appropriate proteins. During carbohydrate metabolism, a prosthetic group consisting of bound lipoic acid (Fig. 11) carries acetyl groups. Both acetyl groups and long-chain fatty acyl groups are carried across membranes into and out of mitochondria while attached to the unusual amino acid carnitine (Fig. 4). Carnitine is not a vitamin but acts as a coenzyme.

B. Nicotinamide Adenine Dinucleotide (NAD^+) and Flavin Adenine Dinucleotide (FAD), Hydrogen and Electron Carriers

Because of the linkage of the vitamin nicotinamide to the ring of the sugar ribose, NAD^+ and its relative $NADP^+$ (which carries an extra phospho group in its structure; Fig. 8) can be reduced by transfer of a hydrogen atom from an alcohol or other suitable substrate to the 4 position of the ring. As illustrated in Fig. 8, the transfer is that of a hydrogen atom plus an electron (a hydride ion H^-). NAD^+ plays this role in many biological dehydrogenation reactions which convert various alcohols into the corresponding carbonyl compounds—aldehydes or ketones. At the same time, many carbonyl compounds are reduced to alcohols. Sometimes the oxidation and reduction processes are linked. A well-known example is the oxidation of glyceraldehyde 3-phosphate during the breakdown of glucose, a process that occurs in bacteria, yeast, and the human body. In all cases NAD^+ is reduced to $NADH + H^+$. The latter is reoxidized to NAD^+ in the human body, but in lactic acid bacteria the NADH is used (always together with an H^+ ion) to reduce pyruvic acid to lactic acid. This provides a balanced fermentation process that requires no oxygen. Under conditions of extreme exertion, e.g., in a 100-meter race, the lactic acid fermentation fuels human muscles. In yeast, a similar fermentation reduces acetaldehyde to ethanol, indirectly providing energy for the cell.

Why are there two similar coenzymes NAD and NADP? A generalization that holds in many instances is that NAD^+ initiates dehydrogenation (oxidation) while NADPH acts as a biological reductant. This permits oxidative pathways utilizing NAD^+ to occur at the same time as reductive processes that utilize NADPH. Cells of aerobic organisms often keep the concentration ratio of the reactants $[NAD^+]/[NADH]$ high at the same time that the ratio $[NADPH] / [NADP^+]$ is also high. Nicotinamide is a very stable compound, but the coenzyme forms are surprisingly easily destroyed. The reduced forms NADH and NADPH are extremely unstable below pH 7, undergoing ring opening reactions. NAD^+ and $NADP^+$ are unstable at high pH, hydroxide ions adding to double bonds in the nicotinamide ring with subsequent destruction of the coenzymes. It is not surprising that our bodies need a daily supply of this vitamin.

Like NAD^+, FAD and the simpler riboflavin monophosphate (FMN) often serve as an acceptor of a hydride (H^-) ion. However, FAD is a more powerful oxidant than is NAD^+. This fact is indicated in a quantitative way by the standard reduction potential, which biochemists tabulate for pH 7. At this pH the standard hydrogen electrode potential $E^{0'}$ (for the couple H^+/H_2) is -0.414 V while that for the powerful oxidant O_2 (O_2/H_2O) is $+0.815$ at 25°C. For the NAD^+/NADH couple $E^{0'}$ is -0.32 V and for FAD/$FADH_2$ it is -0.21 V. However, since FAD and $FADH_2$ are often tightly bound as flavoproteins, the value of $E^{0'}$ for flavoproteins varies over a broad range from -0.49 to $+0.19$ V. The value depends upon the relative strength of binding of the oxidized and reduced forms of FAD to the specific catalytic proteins. In the β oxidation of fatty acids (Fig. 12), the powerful oxidizing properties of FAD make it possible to remove a C3 hydrogen atom as H^- either after or concurrently with the removal of a proton from C2. The latter requires participation of a basic group from the protein as well as activation by the CoA thioester group (step *b* in Fig. 12). The thioester group

also facilitates addition of an HO^- ion at the C3 position in step *c* to form an alcohol. The latter is dehydrogenated by NAD^+ in step *d*.

Another important aspect of FAD chemistry is the ability to accept a single hydrogen atom (or a single electron together with a proton) to form a free radical, which we may designate FADH•; the dot indicates the reactive unpaired electron. This ability allows FAD or FMN to accept a hydride ion, undergoing a two-electron reduction, then pass the electrons one at a time to an electron-accepting metal center in an electron transport chain such as that found in membranes of the mitochondria. It is at the ends of these electron transport chains that oxygen (O_2), brought into the human body through the lungs, combines with four electrons and four protons to form two water molecules. At the other end of the chain —OH groups in a variety of metabolic intermediates are dehydrogenated to carbonyl groups by molecules of NAD^+. The resulting NADH transfers its hydrogen (plus a free H^+) to FMN within the mitochondrial chain. These reactions, which pass electrons through the electron transport chain, account for most of the oxygen utilized in respiration.

The ability to accept single electrons also allows FAD or FMN attached to some enzyme proteins to react directly with O_2, reducing the O_2 to hydrogen peroxide, H_2O_2. The latter has useful functions within cells but may also cause damage. Molecular oxygen (O_2) combined chemically with the reduced riboflavin is also used by hydroxylases of bacteria and plants to introduce —OH groups into a variety of compounds. A peroxide form of FMN, when bound to the correct protein of luminous bacteria, emits visible light.

Living cells contain many other hydrogen and electron carriers. Among them are lipoic acid (Fig. 11), quinones such as vitamin K, ubiquinone and plastoquinone (Fig. 3), and metal centers containing iron, copper, nickel, manganese, and cobalt.

C. Cleaving C—C Bonds with the Help of Coenzymes

The breakdown of fats, sugars, and other foods as well as the synthesis of body constituents depends upon numerous processes of making and breaking chemical bonds. The cutting and forming of C—C bonds is especially challenging. Enzymes can utilize chemical groupings of an acidic or basic character that are present in the amino acids from which the proteins are made. The acidic —COOH, imidazolium (from histidine), and $—NH_3^+$ groups serve as proton donors, and the unprotonated forms of these same groups, as proton acceptors. These groups facilitate cleavage and formation of O—H, N—H, and C—H bonds. Certain C—C bonds, e.g., those that are one atom removed from a carbonyl group, can also be broken by proteins using only their own catalytic acid–base groups. This is illustrated in Fig. 13. The carbonyl group provides an electron-accepting center into which electrons can flow temporarily as the C—C bond is broken. Both the cleavage of a *β* oxoalcohol and a *β* oxoacid (decarboxylation) are illustrated. In some instances the carboxyl (—COOH) group of a substrate can serve as an electron acceptor. However, the reactivity toward bond cleavage is much higher in a thioester such as that formed from acetyl-coenzyme A (Fig. 10). This high reactivity accounts for one function of coenzyme A. For example, coenzyme A permits the cleavage, by a reverse Claisen condensation, of the fatty acid chain during the *β* oxidation of fatty acids (Fig. 12, step *e*). However, not all C—C bonds can be broken using only the chemical groupings of the proteins or of coenzyme A.

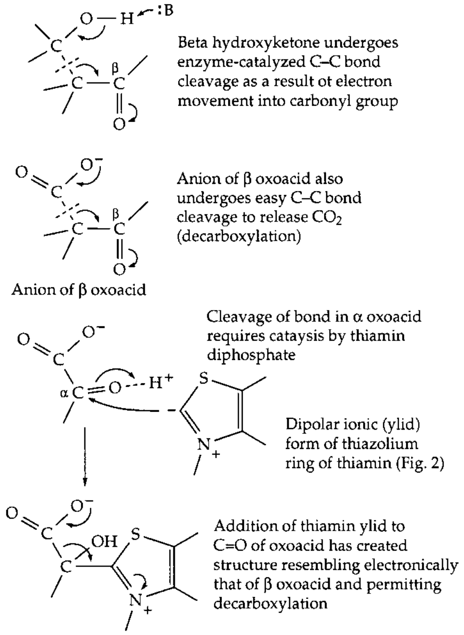


FIGURE 13 Activation of C—C bond cleavage by adjacent carbonyl group (top) and by formation of adduct with thiamin diphosphate (bottom).

Participation of thiamin diphosphate or pyridoxal phosphate is required for many other C—C bond cleavages. Thiamin diphosphate enables cleavage of an *α* oxoacid as indicated in Fig. 13. A characteristic of thiamin diphosphate is that, when bound correctly into an active site, it can lose a proton from its 5-membered thiazolium ring to form the dipolar ionic "ylid" structure shown in Fig. 13. This can add to the carbonyl group of an *α* oxoacid or an *α* oxoalcohol to form a covalent compound (adduct) in which the double bond of the thiazolium ring provides the necessary electron-accepting center. The positive change on the nitrogen atom of the ring assists in initiating the chain cleavage. These thiamin-dependent cleavage

reactions are found in virtually every major metabolic pathway in higher organisms and in most bacteria. For example, the acetyl-CoA that is generated by β oxidation of fatty acids (Fig. 12) enters the citric acid cycle where the two carbon atoms of the acetyl group are converted to CO_2. One essential step in the cycle requires thiamin diphosphate. It is hard to imagine how such metabolic cycles could be organized without thiamin diphosphate.

Pyridoxal phosphate, sometimes in collaboration with pyridoxamine phosphate, participates in dozens of different reactions of amino acids, the building blocks of proteins. These reactions involve both the biological synthesis of amino acids and the breakdown of amino acids, e.g., of excess amino acids in the human diet. For these reactions, the PLP is held in place by the enzyme in a location adjoining the binding site for the specific amino acid substrate. In this site an amino group of a protein side chain (a lysine side chain; Protein —NH_2) forms a Schiff base linkage in which the carbonyl (C=O) group of PLP is converted to a Schiff base linkage (C=N—Protein) similar to that present in the PLP Schiff base drawn in Fig. 14. This is the "resting form" of the enzyme. Then, in a two-step process, the amino group of the substrate adds to the C=N bond and displaces (eliminates) the Protein —NH_2 group to form the substrate Schiff base that is shown in generalized form in Fig. 14. In this Schiff base, one of the three bonds (a, b, c,) may be broken. This is illustrated for cleavage a, removal of a hydrogen atom as H^+ by a catalytic group of the protein. The small arrows beside the structure indicate the manner in which the pyridine ring of

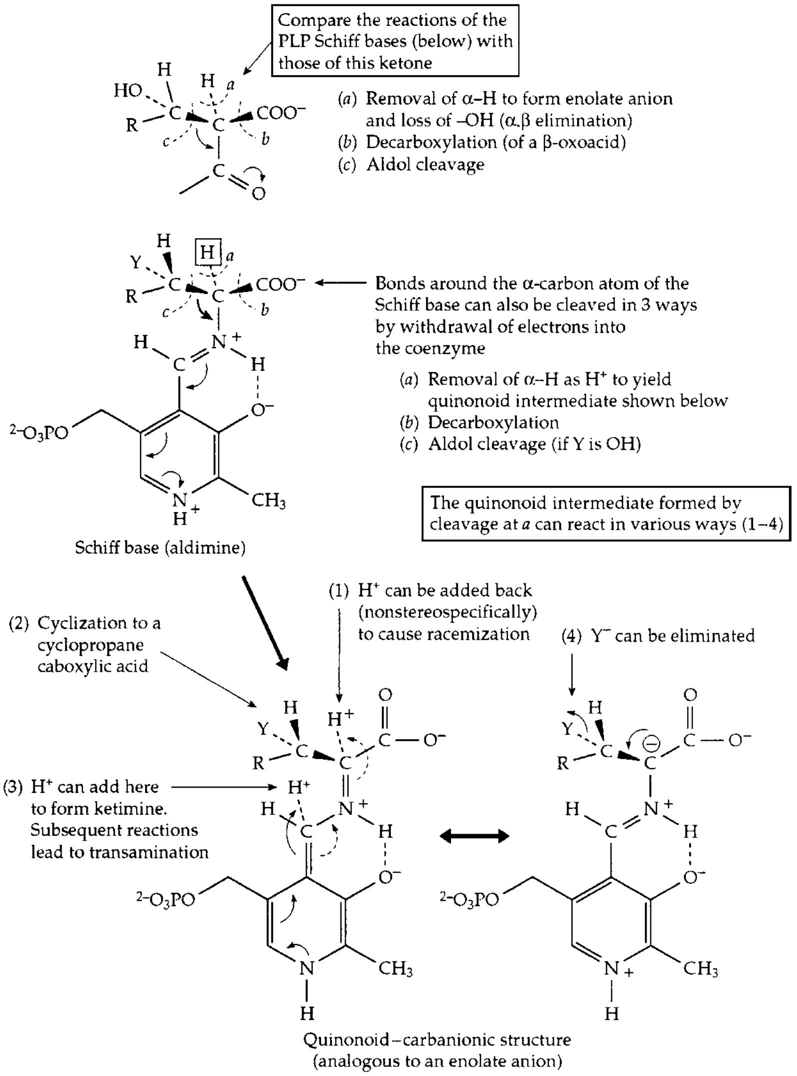


FIGURE 14 The action of pyridoxal phosphate in initiating catalysis of numerous reactions of α-amino acids. Completion of the various reactions requires a large variety of different enzyme proteins.

the coenzyme, with a proton bound to the nitrogen atom of its ring, serves as an electron acceptor in much the same way as does thiamin diphosphate (Fig. 13). The structure resulting from removal of the α-proton of the PLP Schiff base is variously known as a "quinonoid" or "carbanionic" intermediate. Depending upon the specificity of the enzyme in whose site it is formed, this intermediate may react in several ways. In a bacterial racemase a proton may be returned to the α-carbon atom from which it was removed but without stereospecificity, i.e., into either of two positions relative to the other groups surrounding the α-carbon. Some racemases are used by bacteria to convert the stereoisomer known as L-alanine into the less common "unnatural" D-alanine. The latter is incorporated into the bacterial cell wall and helps provide protection to the bacteria against attack by hydrolytic enzymes.

A second mode of reaction of the quinonoid-carbanionic intermediate is utilized by plants which synthesize an enzyme that acts on the amino acid *S*-adenosylmethionine to form a cyclic three-membered ring compound aminocyclopropane carboxylic acid. This is a major plant hormone. In a third type of reaction a proton is added back to the coenzyme itself (see Fig. 14) to form what is called a ketimine (not illustrated). This is a Schiff base of pyridoxamine phosphate (PMP, Fig. 5) with an α-oxoacid and is an essential intermediate compound in the important process of transamination (Fig. 14). This process is utilized by all living organisms both in the synthesis of amino acids and in the breakdown of excesses of amino acids. The human body forms several amino acids via transamination. As shown in Fig. 15, this is a reversible sequence involving a cyclic interconversion of PLP and PMP in reaction steps of the type illustrated in Fig. 14.

Yet another reaction for the ketimine illustrated in Fig. 14 is the elimination of a substituent (labeled Y in this drawing) with formation of a double bond. The product of this elimination sometimes decomposes, with loss of nitrogen as ammonia (NH_3), but in other cases a molecule carrying a different group may replace Y. Protonation of the new Schiff base that results yields a new amino acid. Several amino acids are made by plants and microorganisms using this reaction sequence. Returning to the top of Fig. 14, notice that cleavage of bond *b* leads to formation of CO_2 and decarboxylation of the substrate amino acid. In this way the amino acid dihydroxyphenylalanine (dopa) is converted to the neurotransmitter dopamine. The latter can then be hydroxylated and methylated to form the hormone adrenaline. Histidine is converted by decarboxylation to histamine, a problem compound in allergic reactions, while in the brain, the major excitatory neurotransmitter is decarboxylated to gamma-aminobutyrate (gaba). This is the major inhibitory transmitter in the central nervous system and the compound that keeps our brains calm enough to function. Cleavage of bond *c* (Fig. 14), when R=H and Y=OH (the amino acid is serine) releases the single-carbon compound formaldehyde. This process also requires tetrahydrofolate (Fig. 6). In a converse type of reaction glycine or serine may be condensed with various carbonyl compounds to initiate new biosynthetic pathways. These are often coupled to decarboxylation, which helps to drive the sequence in the biosynthetic direction. One of these yields the red heme pigment of blood.

A third coenzyme that is involved in C—C bond cleavage and formation is the vitamin B_{12} derivative 5′-deoxyadenosylcobalamin (Fig. 7). In this compound the cobalt–carbon bond is easily cleaved to form a free radical which, in turn, facilitates C—C bond cleavage in the substrate. The details, which are still under study, have been omitted, but Fig. 16 shows a general reaction in which

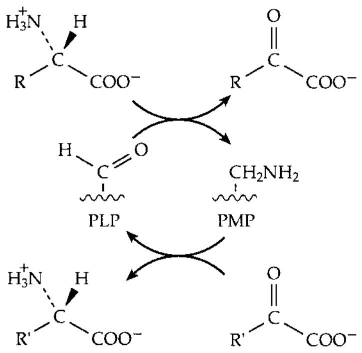


FIGURE 15 The transamination reaction by which amino groups are transferred from one carbon skeleton (in the form of an α oxoacid) to another to form or to degrade an amino acid.

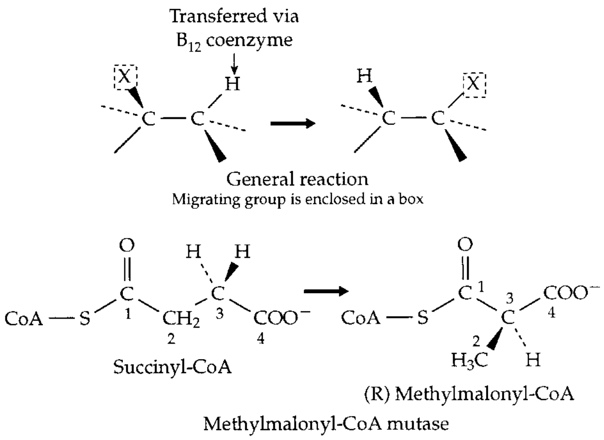


FIGURE 16 (Top) A family of rearrangement reactions that depend upon free radical formation involving an enzyme-bound form of the vitamin B_{12} coenzyme 5′-deoxyadenosylcobalamin (Fig. 7). The rearrangement of (*R*) methylmalonyl-CoA to succinyl-CoA (the opposite of the reaction shown here) is one of the two essential vitamin B_{12}-dependent reactions in the human body, and plays an important role in fatty acid oxidation, as is indicated in Fig. 12.

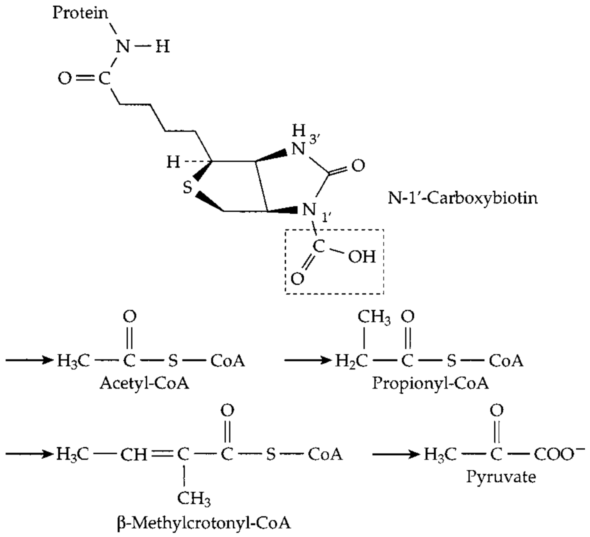


FIGURE 17 The carboxyl carrier function of biotin. A molecule of activated CO_2 is carried as —COOH bonded to N–1′ of biotin, which is covalently attached (as in Fig. 11) to an appropriate protein. Below this structure the sites of four different metabolic intermediates that receive activated CO_2 from carboxybiotin are marked by arrows. In each case, either the thioester linkage to coenzyme A or another adjacent carbonyl group activates a hydrogen atom which dissociates as H^+, leaving a negatively charged site which accepts the CO_2 by direct transfer from carboxybiotin. Carboxylation of propionyl-CoA in the human body is an essential step in degradation of branched chain and odd chain-length fatty acids (Fig. 12). The resulting methylmalonyl-CoA is converted to succinyl-CoA, the reverse of the reaction shown in Fig. 16.

group X is often attached via a C—C bond which is broken. The net result is that a hydrogen atom trades places with group X. These rearrangement reactions, which cannot be catalyzed by proteins alone or by other coenzymes, are quite numerous in various bacteria. However, only one of them occurs in human cells. That is the conversion of methylmalonyl-CoA to succinyl-CoA, the reverse of the succinyl-CoA mutase reaction as drawn in the lower section of Fig. 16. The reaction is essential to the metabolism of propionyl-CoA as is indicated at the bottom of Fig. 12. Propionyl-CoA is carboxylated at the site marked by an arrow in Fig. 17 to form methylmalonyl-CoA. This compound must be isomerized by the vitamin B_{12}-dependent mutase to form succinyl-CoA which can be oxidized to CO_2 in the body's central metabolic pathways. Lack of the mutase is fatal.

D. Carriers of Single-Carbon Compounds, and Other Roles of Pterin Coenzymes

The three coenzymes biotin, tetrahydrofolate, and the vitamin B_{12} derivative methylcobalamin (Fig. 7) act as carriers of the single-carbon compounds CO_2, bicarbonate ions, formaldehyde, and formic acid. The combining of biotin with CO_2 is not a spontaneous process but depends upon adenosine triphosphate (ATP), which serves as both a phospho group carrier and the common energy currency for many cellular reactions. It can also be regarded as a coenzyme. In order to be activated by reaction with ATP, the CO_2 must first combine with a hydroxide ion to form bicarbonate HCO_3^-. ATP then transfers a phospho group to the bicarbonate, forming the labile and short-lived carboxyl phosphate ($^-OOC{-}O{-}PO_3^{2-}$) together with adenosine diphosphate (ADP). The carboxyl phosphate, in turn, transfers the carboxyl group to the biotin prosthetic groups of the various carboxylase proteins. From them the carboxyl group is transferred onto the various sites marked by arrows in Fig. 17. An inorganic phosphate ion is released when the carboxyl group is transferred to biotin, completing a sequence that couples activation of CO_2 with the cleavage of ATP to ADP and inorganic phosphate (P_i). Such coupling of ATP cleavage to biosynthesis is a common feature of much of biosynthetic metabolism.

Two other biotin-dependent reactions of great significance are the carboxylation of acetyl-CoA to malonyl-CoA and that of pyruvate to oxaloacetate (Fig. 17). The former is essential to biosynthesis of fatty acids, which are formed in a pathway which parallels (in reverse) that of β oxidation (Fig. 12). However, there are several differences. In the biosynthetic pathway, acetyl-CoA is first converted to malonyl-CoA which undergoes decarboxylation when a two-carbon unit is added to the growing fatty acid chain. This decarboxylation, together with the prior carboxylation steps, couples ATP cleavage to the biosynthesis. Furthermore, NADPH is used in the reduction steps rather than NADH or $FADH_2$. In addition, the acyl carrier is not coenzyme A but the related prosthetic group of acyl carrier protein. Another biosynthetic process that depends upon biotin is the synthesis of glucose in the liver. Pyruvate, a product of glucose breakdown, is carboxylated to oxaloacetate which is later decarboxylated on its pathway to glucose. Again ATP cleavage is coupled to biosynthesis with the help of biotin.

Tetrahydrofolates (THF) interconvert several one-carbon compounds or fragments. As is indicated in Fig. 18, formaldehyde released from the PLP-dependent cleavage of serine is immediately trapped by THF (Fig. 14). Nitrogen N1 adds to formaldehyde to form a carboxymethyl (—CH_2—COOH) derivative which can than react reversibly with loss of water to form a cyclic adduct (Fig. 18). This compound can be oxidized to the N^{10}-methyl form. Both of these are important intermediates in a variety of biosynthetic processes. The third one-carbon carrier is vitamin B_{12} which can act as an acceptor, taking the methyl group from methyl-THF to form

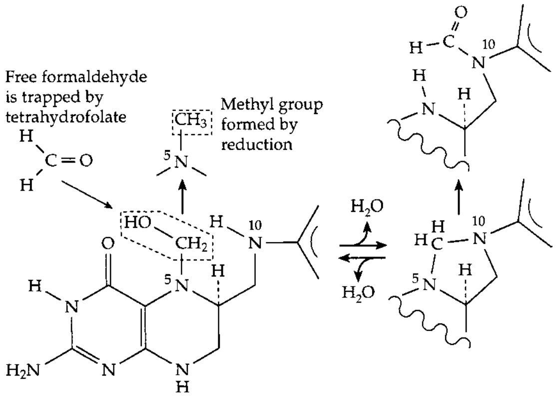


FIGURE 18 The functioning of tetrahydrofolates (THF) in oxidation and reduction of single-carbon fragments. A PLP-dependent enzyme cleaves serine (Fig. 14), releasing formaldehyde, which combines in the active center with THF. Formic acid can be converted to formyl-THF. The various THF derivatives supply single-carbon fragments for many biosynthetic processes.

methylcobalamin (Fig. 7). This compound is transferred to the amino acid homocysteine to form methionine, one of the 20 major amino acids from which proteins are constructed. The reaction accounts for the second human requirement for vitamin B_{12}. If the methionine dietary intake is high enough, this reaction is less important, but the enzyme is still essential for remethylation of homocysteine formed when methionine is used in a variety of processes of biological methylation.

The double ring system on which folic acid (Fig. 6) is constructed is known as pterin. In addition to the folates, a number of other pterin coenzymes are found in the human body and elsewhere in nature. Several have shorter side chains at the 6-position on the ring. Some of these compounds are used to color butterfly wings. Another, called biopterin, has a three-carbon side chain that carries two hydroxyl groups. Its reduced form, tetrahydrobiopterin, is a coenzyme for a series of hydroxylases. Among these is phenylalanine hydroxylase which is lacking in the well-known human genetic defect phenylketonuria (PKU). The reduced pterin ring has properties similar to those of $FADH_2$. Molecular oxygen (O_2) can add to form a peroxide that can donate an OH group (formally as ^+OH) to convert phenylalanine to tyrosine. Phenylalanine is toxic to the brain, accounting for the devastating symptoms of PKU. Another pterin derivative is molybdopterin, which has a four-carbon side chain containing two sulfur atoms and an —OH group. The human body, as well as all other organisms, connects this —OH group to a guanine nucleotide to give a complex cofactor somewhat resembling NAD^+. The two sulfur atoms, however, bind to an atom of the metal molybdenum. The molybdenum atom is the site at which our bodies oxidize the toxic sulfite (SO_3^{2-}) to the harmless sulfate (SO_4^{2-}). This coenzyme, and the associated enzyme sulfite oxidase, are essential to human life. Methane-forming bacteria create a different complex side chain in methanopterin, which replaces tetrahydrofolate in those organisms. Very complex pterin derivatives form the red eye pigments of fruit flies.

E. Antioxidant Systems

Vitamins C and E, as well as ubiquinones (Fig. 3) and derivatives of the nonmetallic element selenium, together with sulfur-containing proteins, all participate in an elaborate antioxidant system. This system protects us against many of the adverse effects of reduced oxygen compounds such as hydrogen peroxide (H_2O_2), superoxide (O_2^-), and hydroxyl (OH) radicals. The system is quite complex and not fully understood. However, ascorbic acid, which can itself form free radicals readily, appears to be a key player. Water-soluble and present in a high concentration, its role seems to be to keep many cellular components reduced. The tocopherols (vitamin E), in their various isomeric forms, scavenge free radicals formed from oxidation of unsaturated fatty acids within cell membranes. Vitamin E is especially effective in removing organic peroxide radicals. Supplementation with dietary vitamin E is being tested for prevention or amelioration of a variety of diseases of aging including atherosclerosis and Parkinson's and Alzheimer's diseases. The resulting tocopherol radicals are rereduced by ascorbate in the aqueous phase. Ascorbate can also donate electrons to ubiquinone radicals present in the membranes of the mitochondria. It is within the mitochondria that many damaging radicals are thought to arise as side products of the reduction of O_2 that occurs there.

In addition to its antioxidant role, ascorbic acid functions to keep various metallic ions in catalytic centers in their reduced forms. For example, some oxygenases require iron or copper in their Fe^{2+} or Cu^+ states of oxidation. If these protein-bound ions are accidentally left in a more oxidized state they may need to be reduced by ascorbate ions. While this is a protectant role, there are some enzymes for which ascorbate has become a cosubstrate. An example is dopamine β-hydroxylase, which converts dopamine to the neurotransmitter noradrenaline. The enzyme contains copper which cycles between Cu^+ and Cu^{2+}, as it incorporates one atom of oxygen from O_2 into its substrate. Ascorbate supplies the electrons for reduction of the second atom of the O_2 to H_2O. A recent report describes another distinct function for ascorbate ion. It apparently acts as a basic catalytic group for proton abstraction from a water molecule during the action of a glycosyltransferase enzyme, becoming part of the active site of that enzyme.

F. Vitamin A and Vision

While vitamin A, as retinoic acid, has important hormonal actions (which are not discussed here), its best known function is in vision. Within photoreceptor cells of the retina, and even in certain bacteria, vitamin A aldehyde (retinal, Fig. 1) forms a Schiff base with specific lysine side chains of the light receptor proteins. Two of the best known of these receptors are rhodopsin, the pigment present in the rod cells of the mammalian retina, and bacteriorhodopsin, the light receptor of the purple membranes of certain salt-tolerant bacteria. In both of these cases, the protein consists of a similar bundle of seven connected helical segments that pass through a membrane. The retinal Schiff base is inside the bundle, held rigidly in a small "box." In both cases, a particular stereoisomer of retinal is present. In bacteriorhodopsin it is the all-*trans* isomer pictured in Fig. 1, but in rhodopsin it is the 11-*cis* isomer shown in Fig. 19. Upon absorption of light, this isomer is converted almost instantaneously into the all-*trans* form as shown in Fig. 19. The all-*trans* retinal then leaves the photoreceptor and is replaced with a new molecule of the 11-*cis* isomer before the photoreceptor can act again. In bacteriorhodopsin, absorption of light converts the all-*trans* retinal into the 13-*cis* isomer within about three trillionths of a second. In both cases, the change in shape of the retinal upon absorption of light induces a small alteration in the geometry and chemical properties of the photoreceptor protein that surrounds the light-absorbing molecule. This is enough to start a chain of signaling events in the retina that leads to a nerve impulse being sent to the brain. In the bacteria, the light absorption is used in a different way to pump a proton from the inside of the cell across the membrane to the outside. The resulting gradient of hydrogen ions (positive charges) across the membrane represents a store of protonic energy similar to that in an electrical condenser. It is used by these cells as a source of energy.

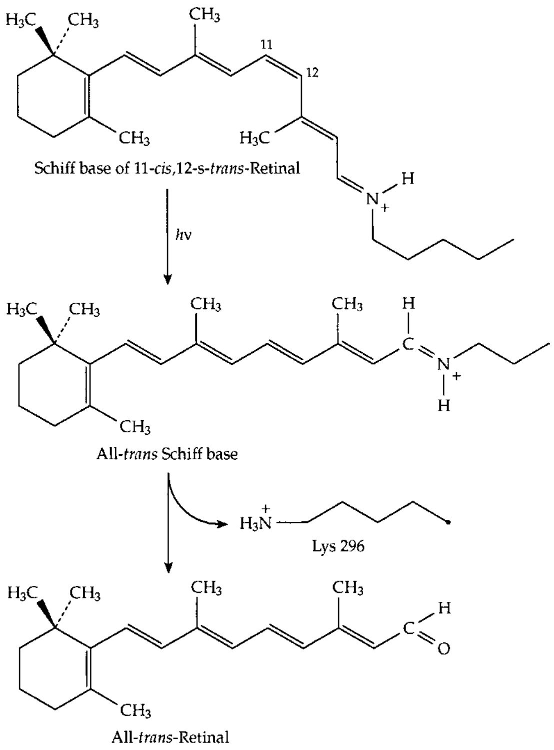


FIGURE 19 The structural change that takes place in the Schiff base of retinal (vitamin A aldehyde) that is formed with specific lysine side chains of the visual pigment proteins upon absorption of a quantum of light. This change triggers a cycle of alterations in the protein that initiates an impulse in the optic nerve.

G. Vitamins A and D as Prohormones

In addition to the coenzyme function of retinal in vision another vitamin A derivative, retinoic acid, is an important hormone with effects on differentiation of cells and tissues. It acts to control transcription of the genetic messages in DNA by binding to specific protein receptors that in turn bind to specific nucleotide sequences of the DNA. The retinoid receptor proteins are a member of the steroid hormone receptor family. Also related to this family are receptors for hydroxylated derivatives of vitamin D.

Vitamin D can be viewed as a prohormone which arises by the action of ultraviolet light in the two-step process pictured in Fig. 20. Irradiation of 7-dehydrocholesterol in the skin can provide adequate amounts of vitamin D_3 (cholecalciferol or calciol). The closely related vitamin D_2 (ergocalciferol) arises from irradiation of the plant sterol ergosterol. This form of the vitamin has been widely used in fortification of milk. However, the natural vitamin D_3 is more active in preventing rickets. The term vitamin D_1 was dropped when it was found to be a mixture of D_2 and D_3. The principal function of vitamin D is in the control of calcium metabolism. This control is exerted by polar, hydroxylated compounds of which the most important is 1α,12-dihydroxyvitamin D_3 (calcitriol). This hormone is distributed to all parts of the body. In cells of the intestinal lining it promotes uptake of calcium ions. It promotes reabsorption of both calcium and phosphate ions in the kidney tubules and increases blood calcium and depositon of calcium ions in bone.

H. Vitamin K and Blood Clotting

Vitamin K (phylloquinone, Fig. 3), the only form of vitamin K found in plants, functions as an electron carrier in the photosynthetic membranes of the chloroplasts. There it serves to carry electrons from the photosystem I receptor in an electron transport chain related to that of mitochondria. The latter utilizes ubiquinone rather than phylloquinone.

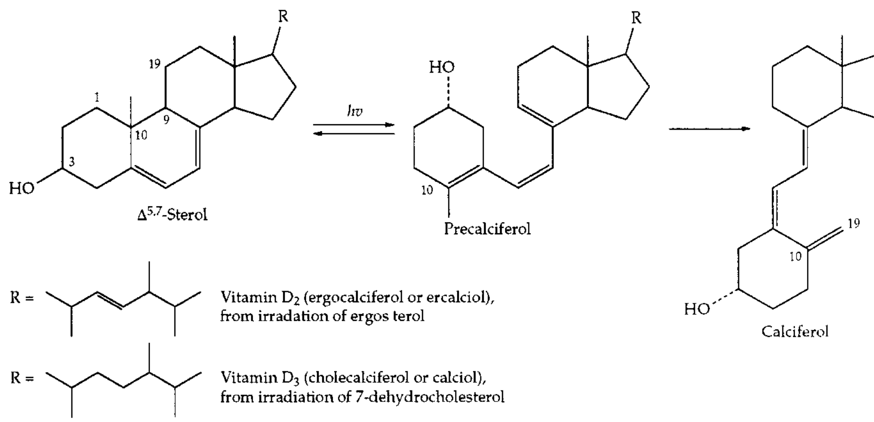


FIGURE 20 Creation of the prohormone vitamin D by the action of light on 7-dehydrocholesterol in human skin or on the plant compound ergosterol.

Some photosynthetic bacteria utilize menaquinone, in which the number of isoprenoid units in the polyprenyl side chain is greater. In the human body vitamin K has a quite different and specialized function in the modification of the side chains of glutamic acid units in a small group of proteins. Among these are prothrombin and other blood clotting proteins. Selected glutamic acid side chains (at 10 positions near the N-terminal end of the prothrombin chain) are modified by addition of an extra carboxyl group at the gamma (or C4) position of the side chain to give γ-carboxyglutamate (Gla) units. The side chain now contains two negatively charged —COO^- groups and is able to better bind to calcium ions (Ca^{2+}), which help bind the clotting factors to the phospholipid membrane in the blood clotting complex. Formation of Gla requires both vitamin K and O_2, as indicated in Fig. 21. Here the quinone form of vitamin K has been reduced to the dihydro form. It has been shown experimentally to be converted to the epoxide derivative as the glutamyl (Glu) side chain is converted to that of Gla. One possible mechanism is depicted in Fig. 22. The O_2 molecule has added to the dihydro-vitamin K to form a peroxide which is used to generate an HO^- ion in the active site where it is in a position to remove the hydrogen in the γ position to form H_2O. The resulting anion adds to CO_2 to form the Gla.

I. Recently Discovered Coenzymes and Prosthetic Groups

While all of the vitamins may have been discovered, new catalytic cofactors and prosthetic groups are still being found. They are too numerous to mention. A few are shown

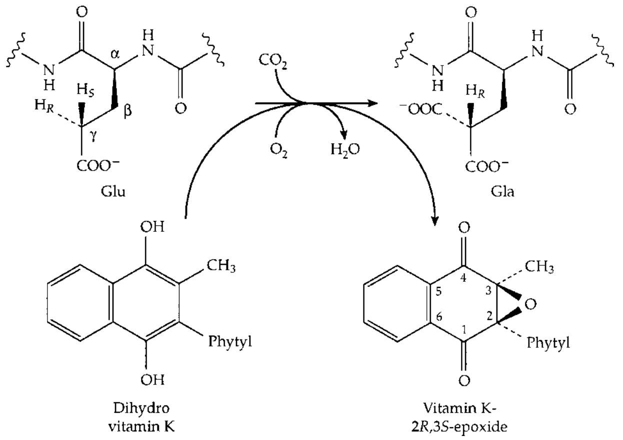


FIGURE 21 Scheme showing the coupling of O_2-dependent oxidation of vitamin K to its epoxide to the carboxylation of the γ-carbon of a glutamyl side chain to a γ-carboxyglutamate (Gla) side chain. One atom of the O_2 enters the epoxide while the other enters H_2O.

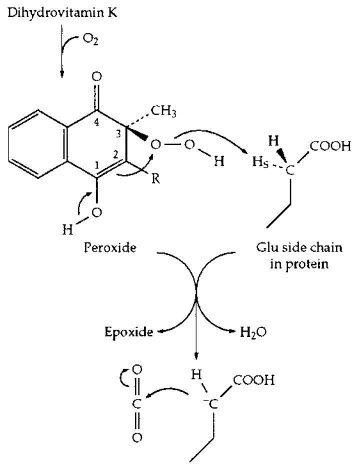


FIGURE 22 A plausible mechanism by which vitamin K acts as a coenzyme to assist in the formation of γ-carboxyglutamate side chains in proteins of the blood-clotting system.

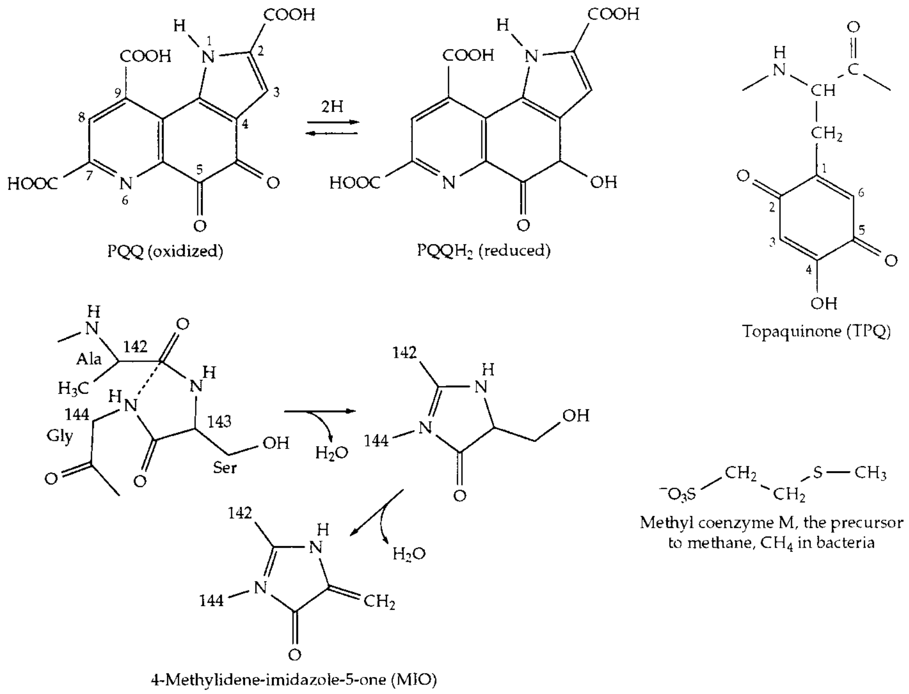


FIGURE 23 Structures of a few recently discovered coenzymes or prosthetic groups.

in Fig. 23. The coenzyme PQQ is a hydrogen carrier that replaces NAD in certain bacterial alcohol dehydrogenases. Topaquinone is a cofactor for amino oxidases found in bacterial, plant, and human tissues. It is a prosthetic group that is formed by oxidation of a specific tyrosine in the active site. It functions together with a nearby copper ion which binds the O_2 substrate. The same copper center and O_2 are apparently involved in the spontaneous synthesis of the prosthetic group. Formation of topaquinone is only one case of many, in which prosthetic groups are self-assembled using components of the protein that carries the group. Another recently discovered example is formation of 4-methylidene-imidazole-5-one (MIO), from glycine and serine units of a protein. The prosthetic group is involved in a previously hard-to-explain isomerization of histidine and phenylalanine. A parallel reaction with phenylalanine initiates the major pathway of synthesis of thousands of aromatic compounds in plants.

Among specialized coenzymes are a collection of compounds (which include the previously mentioned methanopterin) needed by methane-forming bacteria and compounds such as the light-emitting luciferin of fireflies. Luminous jellyfish make different light-emitting compounds which, like topaquinone and MIO, are made from side chains of amino acids. The list could be continued for pages.

SEE ALSO THE FOLLOWING ARTICLES

Enzyme Mechanisms • Natural Antioxidants in Foods • Pharmaceuticals • Protein Structure • Protein Synthesis

BIBLIOGRAPHY

Metzler, D. E. (2001). *Biochemistry, The Chemical Reactions of Living Cells*, 2nd ed. Vol. 1, Harcourt/Academic Press, San Diego.

Shils, E., Olson, J. A., Shike, M., and Ross, A. C., eds. (1999). *Modern Nutrition in Health and Disease*, 9th ed., Williams & Wilkins, Baltimore.

McCormick, D. B., and Chen, H. (1999). *J. Nutrition* **129,** 325–327.

Iriarte, A., Kagan, H. M., Martinez-Carrion, M., eds. (2000). *Biochemistry and Molecular Biology of Vitamin B_6 and PQQ-Dependent Proteins*, Birkhäuser Verlag, Basel, Switzerland.